A Bayesian nonparametric model for count functional data

Antonio Canale and David B. Dunson

Abstract Count functional data arise in a variety of applications, including longitudinal, spatial and imaging studies measuring functional count responses for each subject under study. The literature on statistical models for dependent count data is dominated by models built from hierarchical Poisson components. The Poisson assumption is not warranted in many applications, and hierarchical Poisson models make restrictive assumptions about over-dispersion in marginal distributions. This article discuss a class of nonparametric Bayes count functional data models introduced in Canale and Dunson [3], which are constructed through rounding real-valued underlying processes. Computational algorithms are developed using Markov chain Monte Carlo and the methods are illustrated through application to asthma inhaler usage.

Key words: Generalized linear mixed model; Hierarchical model; Longitudinal data; Splines; Stochastic process.

1 Introduction

A stochastic process $y = \{y(s), s \in \mathscr{D}\}$ is a collection of random variables indexed by $s \in \mathscr{D}$, with the domain $\mathscr{D}$ commonly corresponding to a set of times or spatial locations and $y(s)$ to a random variable observed at a specific time or location s. There is a rich frequentist and Bayesian literature on stochastic processes, with common choices including Gaussian processes and Lévy processes, such as the Poisson, Wiener, beta or gamma process. Gaussian processes provide a convenient and well

Antonio Canale
University of Turin and Collegio Carlo Alberto, Turin, Italy e-mail: antonio.canale@unito.it

David B. Dunson
Duke University, Durham, NC e-mail: dunson@duke.edu

studied choice when $y : \mathscr{D} \to \mathfrak{R}$ is a continuous function. Our interest focuses on the case in which $y : \mathscr{D} \to \mathscr{N} = \{0, \dots, \infty\}$, so that y is a count-valued stochastic process over the domain $\mathscr{D}$. There are many applications of such processes including developemental toxicity epidemiology studies monitoring a count health response over time.

Although there is a rich literature on count stochastic process models for longitudinal and spatial data, most models rely on Poisson hierarchical specifications. Although such models have a flexible mean structure, the Poisson assumption is restrictive in limiting the variance to be equal to the mean, with over-dispersion introduced in marginalizing out the latent processes. Such modeling frameworks have several disadvantages. Firstly the dependence structure is confounded with marginals overdispersion and secondly under-dispersed count data are not accomodate. To relax usual Poisson parametric assumptions [10] exploited a hierarchical specification of the Faddy model [6]. Although the gain in flexibility, the computation for this model is challenging.

In considering models that separate the marginal distribution from the dependence structure, it is natural to focus on copulas. Nikoloulopoulos and Karlis [15] proposed a copula model for bivariate counts that incorporates covariates into the marginal model. Erhard and Czado [5] proposed a copula model for high-dimensional counts, which can potentially allow under-dispersion in the marginals via a Faddy or Conway-Maxwell-Poisson [16] model. Genest and Neslehova [8] provide a review of copula models for counts.

An alternative approach relies on rounding of a stochastic process. For classification it is common to threshold Gaussian process regression [4, 9]. For example, [12] rounded a real discrete autoregressive process to induce an integer-valued time series while [2] used rounding of continuous kernel mixture models to induce nonparametric models for count distributions. In this article we discuss a class of stochastic processes introduced in [3] that map a real-valued stochastic process $y^* : \mathscr{D} \to \mathfrak{R}$ to a count stochastic process $y : \mathscr{D} \to \mathscr{N}$.

2 Rounded Stochastic Processes

2.1 Notation and model formulation

Let $y \in \mathscr{C}$ denote a count-valued stochastic process, with $\mathscr{D} \subset \mathfrak{R}^p$ compact and $\mathscr{C}$ the set of all $\mathscr{D} \to \mathscr{N}$ step functions with unit step and a finite number of jumps in $\mathscr{D}$. Such an assumption is a count process version of the continuity condition routinely assumed for $\mathscr{D} \to \mathfrak{R}$ functions. It ensures that for sufficiently small changes in the input the corresponding change in the output is small, being either zero or one. We are particularly motivated by applications in which counts do not change erratically at nearby times but maintain some degree of similarity.

We choose a prior $y \sim \Pi$, where Π is a probability measure over $(\mathscr{C}, \mathscr{B})$, with $\mathscr{B}(\mathscr{C})$ the Borel σ-algebra of subsets of $\mathscr{C}$. The measure Π induces the marginal probability mass functions

$$\text{pr}\{y(s) = j\} = \Pi\{y : y(s) = j\} = \pi_j(s), \quad j \in \mathscr{N}, \quad s \in \mathscr{D}, \tag{1}$$

and the joint probability mass functions

$$\text{pr}\{y(s_1) = j_1, ..., y(s_k) = j_k\} = \Pi\{y : y(s_1) = j_1, ..., y(s_k) = j_k\} = \pi_{j_1 \cdots j_k}(s_1, ..., s_k) \tag{2}$$

for $j_h \in \mathscr{N}$ and $s_h \in \mathscr{D}$, $h = 1, \ldots, k$, and any $k \geq 1$.

In introducing the Dirichlet process, [7] mentioned three appealing characteristics for nonparametric Bayes priors including large support, interpretability and ease of computation. Our goal is to specify a prior Π that gets as close to this ideal as possible. Starting with large support, we would like to choose a Π that allocates positive probability to arbitrarily small neighborhoods around any $y_0 \in \mathscr{C}$ with respect to an appropriate distance metric, such as L^1. To our knowledge, there is no previously defined stochastic process that satisfies this large support condition. In the absence of prior knowledge that allows one to assume y belongs to a pre-specified subset of $\mathscr{C}$ with probability one, priors must satisfy the large support property to be coherently Bayesian. Large support is also a necessary condition for the posterior for y to concentrate in small neighborhoods of any true $y_0 \in \mathscr{C}$.

With this in mind, we propose to induce a prior $y \sim \Pi$ through

$$y = h(y^*), \quad y^* \sim \Pi^*, \tag{3}$$

where $y^* : \mathscr{D} \to \Re$ is a real-valued stochastic process, h is a thresholding operator from $\mathscr{Y} \to \mathscr{C}$, $\mathscr{Y}$ is the set of all $\mathscr{D} \to \Re$ continuous functions, Π^* is a probability measure over $(\mathscr{Y}, \mathscr{B})$ with $\mathscr{B}(\mathscr{Y})$ Borel sets. Unlike count-valued stochastic processes, there is a rich literature on real-valued stochastic processes. For example, Π^* could be chosen to correspond to a Gaussian process or could be induced through various basis or kernel expansions of y^*.

There are various ways in which the thresholding operator h can be defined. For interpretability and simplicity, it is appealing to maintain similarity between y^* and y in applying h, while restricting $y \in \mathscr{C}$. Hence, using the informal definition of rounding as an operation that reduces the number of digits while keeping the values similar, we focus on a rounding operator that let $y(s) = 0$ if $y^*(s) < 0$ and $y(s) = j$ if $j - 1 \leq y^*(s) < j$ for $j = 1, \ldots, \infty$. Negative values will be mapped to zero, which is the closest non-negative integer, while positive values will be rounded up to the nearest integer. This type of restricted rounding ensures $y(s)$ is a non-negative integer. Using a fixed rounding function h in (3), we rely on flexibility of the prior $y^* \sim \Pi^*$ to induce a flexible prior $y \sim \Pi$. For notational convenience and generality, we let $y(s) = j$ if $y^*(s) \in A_j = [a_j, a_{j+1})$, with $a_0 < \cdots < a_\infty$ and we focus on $a_0 = -\infty, a_j = j - 1, j = 1, \ldots, \infty$.

In certain applications, count data can be naturally viewed as arising through integer-valued rounding of an underlying continuous process. For example, in the

longitudinal tumor count studies of Section 3.1, it tends to be difficult to distinguish individual tumors and it is natural to posit a continuous time-varying tumor burden, with tumors fusing together and falling off over time. In collecting the data, tumor biologists attempt to make an accurate count but even at the same time counts can vary. It is natural to accommodate this with a smoothly-varying continuous tumor burden specific to each animal with measurement errors and rounding producing the observed tumor counts. However, even when there is no clear applied context motivating the existence of an underlying continuous process, the proposed formulation nonetheless leads to a highly flexible and computationally convenient model.

2.2 Count functional data

We have focused on the case in which there is a single count process y observed at locations $s = (s_1, \ldots, s_n)^T$. In many applications, there are instead multiple related count processes $\{y_i, i = 1, \ldots, n\}$, with the ith process observed at locations $s_i = (s_{i1}, \ldots, s_{in_i})^T$. We refer to such data as count functional data. As in other functional data settings, it is of interest to borrow information across the individual functions through use of a hierarchical model. This can be accomplished within our rounded stochastic processes framework by first defining a functional data model for a collection of underlying continuous functions $\{y_i^*, i = 1, \ldots, n\}$, and then letting $y_i = h(y_i^*)$, for $i = 1, \ldots, n$. There is a rich literature on appropriate models for $\{y_i^*, i = 1, \ldots, n\}$ ranging from hierarchical Gaussian processes [1] to wavelet-based functional mixed models [14].

Let $y_i(s)$ denote the count for subject i at time s, $y_{it} = y_i(s_{it})$ the number at the tth observation time, and x_{it}a predictor for subject i at time t. As a simple model motivated by the asthma inhaler use applications described below, we let

$$y_{it} = h(y_{it}^*), \quad y_{it}^* = \xi_i + b(x_{it})^T \theta + \varepsilon_{it}, \quad \xi_i \sim Q, \quad \varepsilon_{it} \sim N(0, \tau^{-1}), \tag{4}$$

where ξ_i is a subject-specific random effect, $b(\cdot)$ are B-splines basis functions that depend on predictors and time, θ are unknown basis coefficients, and ε_{it} is a residual. To induce a penalization on finite differences of the coefficients of adjacent B-spline we let $p(\theta \mid \lambda) \propto \exp(-1/2\lambda\, \theta^{\mathrm{T}} P \theta)$, where $P = D^{\mathrm{T}} D$ is a penalty matrix with D the rth order difference matrix and $\lambda \sim \mathrm{Ga}(\nu/2, \delta\nu/2)$, $\delta \sim \mathrm{Ga}(a, b)$. Such a prior for the basis coefficients induces a penalty on finite differences of the coefficients of adjacent B-splines with the parameter λ being a roughness penalty. Such a construction is known as Bayesian P-spline (penalized B-spline) model [13]. The hyperparameter δ controls dispersion of the prior. By choosing a hyperprior with small a, b values, one induces a prior with heavy tails and good performance in a variety of settings [11]. We additionally choose a hyperprior for the residual precision $p(\tau) \propto \tau^{-1}$. To allow the random effect distribution to be unknown, we choose a Dirichlet process prior, with $Q \sim \mathrm{DP}(\alpha Q_0)$, with α a precision parameter and the base measure Q_0 chosen as $N(0, \psi)$. As commonly done we fix $\alpha = 1$.

3 Asthma inhaler use applications

We analyze data on daily usage of albuterol asthma inhalers [10]. Daily counts of inhaler use were recorded for a period between 36 and 122 days at the Kunsberg School at National Jewish Health in Denver, Colorado for 48 students previously diagnosed with asthma. The total number of observations was 5209. As discussed by Grunwald and coauthors [10], the data are under-dispersed.

Let y_{it} denote the number of times the ith student used the inhaler on day t. Interest focuses on the impact of morning levels of PM_{25}, small particles less than 25 mm in diameter in air pollution, on asthma inhaler use. At each day t, a vector $x_t = (x_{t1}, \ldots, x_{tp})^T$ of environmental variables are recorded including PM_{25}, average daily temperature (Fahrenheit degree/100), % humidity and barometric pressure (mmHg/1000). We modify (4) to include multiple predictors with an additive model structure as follows.

$$y_{it} = h(y^*_{it}), \quad y^*_{it} = \xi_i + \sum_{j=1}^{4} b_j(x_{jt})^T \theta_j + \varepsilon_{it}, \tag{5}$$

where ξ_i is a random effect modeled as described in previous section, b_j is a B-spline basis with θ_j the basis coefficients relative to jth predictor and $\varepsilon_i \sim N(0, \tau^{-1}R)$, with R an AR-1 tridiagonal correlation matrix with correlation parameter ρ. The prior for each θ_j is identical to the prior described above leading to an additive Bayesian P-splines model. Each predictor is normalized to have mean zero and unit variance prior to analysis. The correlation parameter is given a uniform prior on $[-1, 1]$. Computational details are reported in [3].

We ran our Markov chain Monte Carlo algorithm for 10,000 iterations with a 1,000 iteration burn-in discarded. To obtain interpretable summaries of the non-linear covariate effects on the inhaler use counts, we recorded for each predictor at a dense grid of x_{jt} values at each sample after burn-in the conditional expectation of the count for a typical student having $\xi_i = 0$,

$$\begin{aligned} \mu_j(x_{jt}) &= \mathrm{E}(y_{it} \,|\, x_{jt}, x_{j't} = 0, j' \neq j, \xi_i = 0, \theta, \tau, \rho) \\ &\approx \sum_{k=0}^{K} k[\Phi\{a_{k+1}; \mu^*_j(x_{jt}), \tau\} - \Phi\{a_k; \mu^*_j(x_{jt}), \tau\}], \end{aligned} \tag{6}$$

where $\Phi(\cdot; \mu, \tau)$ is the cumulative distribution function of a normal random variable with mean μ and precision τ, K is the 99·99% quantile of $N\{\mu^*_j(x_{jt}), \tau^{-1}\}$, and

$$\mu^*_j(x_{jt}) = b_j(x_{jt})^{\mathrm{T}} \theta_j + \sum_{l \neq j} b_l(0)^{\mathrm{T}} \theta_l, \tag{7}$$

with the other predictors fixed at their mean value. Based on these samples, we calculated posterior means and pointwise 95% credible intervals, with the results reported in Figure 1. Interestingly, each of the predictors had a non-linear impact

on the frequency of inhaler use, with inhaler use increasing with morning levels of PM_{25}.

Previous analysis conducted in [10] tackle the problem under a generalized linear mixed models setup with the Faddy distribution. The mean for each subject i at time t was

$$\mu_{it} = \exp(x_{1t}\beta_1 + \cdots + x_{pt}\beta_p + u_i + e_{it}) \tag{8}$$

where u_i is a subject specific random effect and e_{it} an error modeled as an AR-1 process. They estimated a coefficient of just 0·013 for PM_{25}, which is close to zero with 95% intervals including zero. In contrast, we obtain clear evidence of non-linear effects of several of the covariates including PM_{25}.

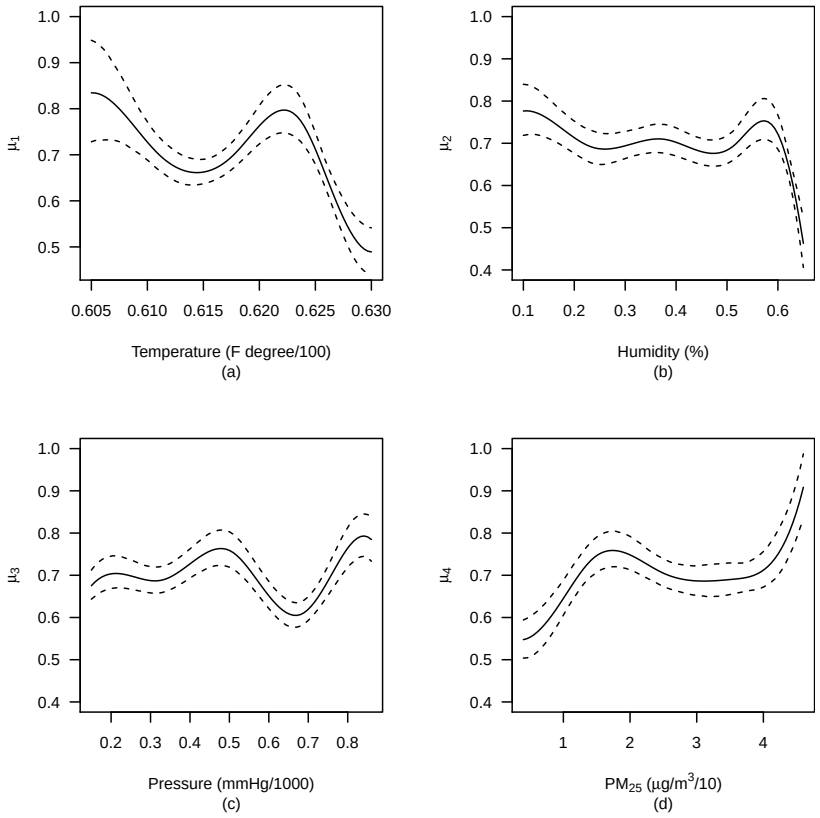

Fig. 1 Posterior mean and 95% pointwise credible bands for the effect of (a) average daily temperature, (b) % of humidity, (c) barometric pressure, and (d) concentration of PM_{25} pollutant on asthma inhaler use calculated with equation (6).

4 Discussion

We have discussed a simple approach, introduced by [3] for modeling count stochastic processes based on rounding continuous stochastic processes. The general strategy is flexible and allows one to leverage existing algorithms and code for posterior computation for continuous stochastic processes. Although rounding of continuous underlying processes is quite common for binary and categorical data, such approaches have not to our knowledge been applied to induce new families of count stochastic processes. Instead, the vast majority of the literature for count processes relies on Poisson process and hierarchical Poisson constructions, which have some well known limitations in terms of flexibility. The modeling framework can be easily generalized to the settings of count functional data, i.e. when one observe n different realizations of a stochastic process and its performance has been shown in an application to asthma inhaler use.

Acknowledgements

This research was partially supported by grant number R01 ES017240-01 from the National Institute of Environmental Health Sciences (NIEHS) of the National Institutes of Health (NIH) and grant CPDA097208/09 by University of Padua.

References

1. Behseta, S., Kass, R.E., Wallstrom, G.L.: Hierarchical models for assessing variability among functions. Biometrika **92**(2), 419–434 (2005)
2. Canale, A., Dunson, D.B.: Bayesian kernel mixtures for counts. Journal of the American Statistical Association **106**(496), 1528–1539 (2011)
3. Canale, A., Dunson, D.B.: Nonparametric Bayes modeling of count processes (2012). Submitted
4. Chu, W., Ghahramani, Z.: Gaussian process for ordinal regression. Journal of Machine learning Research **6**, 1019–1041 (2005)
5. Erhard, V., Czado, C.: Sampling count variables with specified Pearson correlation - a comparison between a naive and a C-vine sampling approach. In: D. Kurowicka, H. Joe (eds.) Dependence Modeling - Handbook on Vine Copulae, pp. 73–87. World Scientific (2009)
6. Faddy, M.J.: Extended Poisson process modeling and analysis of count data. Biometrical Journal **39**, 431–440 (1997)
7. Ferguson, T.S.: A Bayesian analysis of some nonparametric problems. The Annals of Statistics **1**, 209–230 (1973)
8. Genest, C., Neslehova, J.: A primer on copulas for count data. Astin Bulletin **37**, 475–515 (2007)
9. Ghosal, S., Roy, A.: Posterior consistency of Gaussian process prior for nonparametric binary regression. The Annals of Statistics **34**(5), 2413–2429 (2006)
10. Grunwald, G.K., Bruce, S.L., Jiang, L., Strand, M., Rabinovitch, N.: A statistical model for under- or overdispersed clustered and longitudinal count data. Biometrical Journal **53**(4), 578–594 (2011).

11. Jullion, A., Lambert, P.: Robust specification of the roughness penalty prior distribution in spatially adaptive Bayesian P-splines models. Computational Statistics and Data Analysis **51**(5), 2542–2558 (2007)
12. Kachour, M., Yao, J.F.: First order rounded integer-valued autoregressive (RINAR(1)) process. Journal of time series analysis **30**(4), 417–448 (2009)
13. Lang, S., Brezger, A.: Bayesian P-splines. Journal of Computational and Graphical Statistics **13**, 183–212 (2004)
14. Morris, J., Carroll, R.: Wavelet-based functional mixed models. JRSS-B **68**, 179–199 (2006)
15. Nikoloulopoulos, A., Karlis, D.: Regression in a copula model for bivariate count data. Journal of Applied Statistics **37**(9), 1555–1568 (2010)
16. Shmueli, G., Minka, T.P., Kadane, J.B., Borle, S., Boatwright, P.: A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. JRSS-C **54**(1), 127–142 (2005).