Nonparametric Bayes modelling of count processes

Nonparametric Bayes modelling of count processes

Biometrika, 100, 4, 2013, doi: 10.1093/biomet/ast037

Biometrika (2013), *pp*. 1–16

Nonparametric Bayes modelling of count processes

BY ANTONIO CANALE

Department of Economics and Statistics, University of Torino, Corso Unione Sovietica 218/bis, 10134 Torino, Italy

antonio.canale@unito.it

DAVID B. DUNSON

Department of Statistical Science, Duke University, Box 90251, Durham, North Carolina 27708-0251, USA

dunson@stat.duke.edu

SUMMARY

Data on count processes arise in a variety of applications, including longitudinal, spatial and imaging studies measuring count responses. 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 proposes a class of nonparametric Bayes count process models, which are constructed through rounding real-valued underlying processes. The proposed class of models accommodates applications in which one observes separate count-valued functional data for each subject under study. Theoretical results on large support and posterior consistency are established, and computational algorithms are developed using Markov chain Monte Carlo. The methods are evaluated via simulation studies and illustrated through application to longitudinal tumor counts and asthma inhaler usage.

Some key words: Count functional data; Generalized linear mixed model; Hierarchical model; Longitudinal data; Poisson; Spline; Stochastic process.

1. INTRODUCTION

A stochastic process $y = \{y(s), s \in \mathcal{S}\}$ is a collection of random variables indexed by $s \in \mathcal{S}$, with the domain $\mathcal{S}$ 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 studied choice when $y : \mathcal{S} \to \Re$ is a continuous function. In the Bayesian literature, there have been substantial computational and theoretical advances for Gaussian process models in recent years. For example, Banerjee et al. (2008) and Murray & Adams (2010) develop improved methods for posterior computation, while Ghosal & Roy (2006) and van der Vaart & van Zanten (2009) study asymptotic properties including posterior consistency and rates of convergence. The Gaussian process is appealing in providing a prior that can be specified to generate functions that are within an arbitrarily small neighborhood of any continuous function with positive probability (Ghosal & Roy, 2006), while also being computationally convenient.

Our interest focuses on the case in which $y : \mathcal{S} \to \mathcal{N} = \{0, \ldots, \infty\}$, so that y is a count-valued stochastic process over the domain $\mathcal{S}$. There are many applications of such processes including epidemiology studies monitoring a count biomarker or health response over time for patients and ecology studies recording the number of birds of a given species observed at different locations. Although there is a rich literature on count stochastic process models for longitudinal and spatial data, most models rely on Poisson hierarchical specifications. For example, Frühwirth-Schnatter & Wagner (2006) consider $y(s) \sim \text{Poisson}\{\lambda(s)\}$ with the Poisson mean $\lambda(s)$ varying over time according to a latent process. Rue et al. (2009) recently developed an integrated nested Laplace approximation to the posterior for a broad class of latent Gaussian structured additive regression models. The observed variables are assumed to belong to an exponential family, with the means given an additive model having Gaussian and Gaussian process priors on the unknown components.

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. This leads to a confounding of the dependence structure with the degree of overdispersion in the marginals, in that both are induced through the latent process. Such modeling frameworks cannot accommodate correlated count data that are under-dispersed, and substantial bias can result for non-Poisson over-dispersed data. Relying on a hierarchical Faddy model (Faddy, 1997), Grunwald et al. (2011) developed methods that generalize the Poisson distribution to accommodate under- and over-dispersed longitudinal counts. The Faddy distribution allows the current rate of occurrence to depend on the number of events in a previous interval, and when a dispersion parameter is less than zero the rate decreases with each new event, causing under-dispersion. This is a restrictive type of negative feedback dependence and computation is challenging, taking several days to implement a single analysis.

In considering models that separate the marginal distribution from the dependence structure, it is natural to focus on copulas. Nikoloulopoulos & Karlis (2010) proposed a copula model for bivariate counts that incorporates covariates into the marginal model. Erhard & Czado (2009) proposed a copula model for high-dimensional counts, which can potentially allow under-dispersion in the marginals via a Faddy or Conway–Maxwell–Poisson (Shmueli et al., 2005) model. Genest & Neslehova (2007) provide a review of copula models for counts. To our knowledge, copula models have not yet been developed that are directly applicable to count stochastic processes. Wilson & Ghahramani (2010) proposed a Gaussian copula process model to characterize dependence between arbitrarily many random variables independently of their marginals. Rodríguez et al. (2010) proposed a latent stick-breaking process, which is a nonparametric Bayes approach for a stochastic process with an unknown common marginal distribution modeled via a stick-breaking prior. They considered a spatial count process application, with marginal modeled via a mixture of Poisson distributions and the spatial dependence characterized through a latent Gaussian process. This separates the marginal and dependence structures, but the marginal model is restrictive in being characterized as a mixture of Poisson distributions, computation is intensive, and count functional data are not accommodated.

An alternative approach relies on rounding of a stochastic process. For classification it is common to threshold Gaussian process regression (Chu & Ghahramani, 2005; Ghosal & Roy, 2006). Kachour & Yao (2009) rounded a real discrete autoregressive process to induce an integer-valued time series. Canale & Dunson (2011) used rounding of continuous kernel mixture models to induce nonparametric models for count distributions. This article instead proposes a class of stochastic processes that map a real-valued stochastic process $y^* : \mathcal{S} \to \Re$ to a count stochastic process $y : \mathcal{S} \to \mathcal{N}$.

2. ROUNDED STOCHASTIC PROCESSES

2·1. *Notation and model formulation*

Let $y \in \mathcal{C}$ denote a count-valued stochastic process, with $\mathcal{S} \subset \Re^p$ compact and $\mathcal{C}$ the set of all $\mathcal{S} \to \mathcal{N}$ functions satisfying Assumption 1.

Assumption 1. The stochastic process $y : \mathcal{S} \to \mathcal{N}$ is piecewise constant such that $\mathcal{S} = \bigcup_{l=1}^{L} \mathcal{S}_l(y)$, with $y(s)$ constant within the interior of each set $\mathcal{S}_l(y)$ and with unit increments at the boundaries $\mathcal{B}(y)$. The boundary points fall within the set having the higher value of $y(s)$.

Assumption 1 ensures that for sufficiently small changes in the input the corresponding change in the output is small. We are particularly motivated by applications in which counts do not change erratically at nearby times but maintain some degree of similarity. However, Assumption 1 does not rule out rapidly changing count processes, as one can have arbitrarily many jumps in a tiny interval and still satisfy the assumption. In addition, Assumption 1 is easily relaxed.

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

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

and the joint probability mass functions

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

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

In introducing the Dirichlet process, Ferguson (1973) 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 $y_0 \in \mathcal{C}$ 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 $\mathcal{C}$ 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 $y_0 \in \mathcal{C}$.

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

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

where $y^* : \mathcal{S} \to \Re$ is a real-valued stochastic process, h is a thresholding operator from $\mathcal{Y} \to \mathcal{C}$, $\mathcal{Y}$ is the set of all $\mathcal{S} \to \Re$ continuous functions, and Π^* is a probability measure over $(\mathcal{Y}, \mathcal{B})$ with $\mathcal{B}(\mathcal{Y})$ 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 $y \in \mathcal{C}$. Hence 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

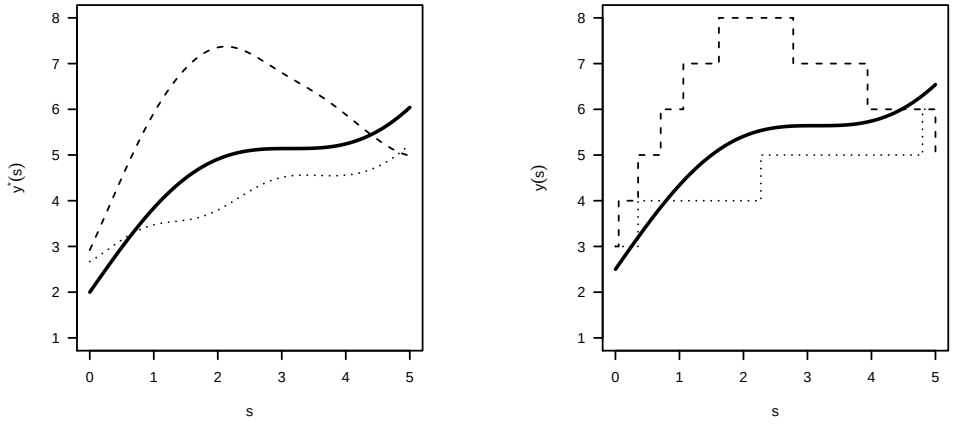

Fig. 1. Thresholding operator h. Panel (a) represents samples from a Gaussian process with mean function $\mu(s) = 2 + \sin(s) + s$ (bold line) and squared exponential covariance function. Panel (b) shows how the mapping operator (3) works. Dotted and dashed lines are the rounded version of panel (a) realizations while the bold line is the induced mean function.

$y \sim \Pi$. 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 $a_0 = -\infty, a_j = j - 1, j = 1, \ldots, \infty$.

This construction is particularly suitable for modeling dynamics of count processes close to zero and in particular, zero-inflated processes with local dependence in the zeros. Applying the mapping h to a latent y^* that assumes negative values across certain sub-regions of $\mathcal{S}$ will lead to blocks of zeros in the count process y. This incorporates dependence between zero occurrences and the occurrence of small counts, which seems natural in most applications such as in the longitudinal tumor count study of §4·2.

Figure 1 illustrates the prior through showing realizations of the underlying stochastic process, in Panel (a), and resulting count process after applying the rounding operator, in Panel (b). The thick lines represents the mean functions of the real valued process and of the induced process. The latter is

$$E\{y(s)\} = \sum_{j=0}^{\infty} j\{F_s(a_{j+1}) - F_s(a_j)\},$$

where $F_s(x) = \int_{-\infty}^{x} f_s(y^*)\mathrm{d}y^*$ and f_s is the marginal distribution of $y^*(s)$.

The covariance structure of the induced count process inherits much of the structure of the underlying process, as is clear from

$$\mathrm{cov}\{y(s), y(s')\} = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} j\,k\,\mathrm{pr}[\{y^*(s) \in A_j, y^*(s') \in A_k\}] - E\{y(s)\}E\{y(s')\},$$

where $\{y^*(s), y^*(s)\}$ has a bivariate distribution with covariance equal to $\mathrm{cov}\{y^*(s), y^*(s)\}$. We report some plots comparing the covariance of the original process with that of the induced process in the Supplementary Material.

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 §4·2, 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 measurement errors are unavoidable. 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. *Properties*

The mapping function $h(\cdot)$ in (3) is many-to-one and the inverse mapping $h^{-1}(y)$ will correspond to an uncountable set of infinitely many continuous stochastic processes y^* such that $y = h(y^*)$. As an important step in characterizing the support of the induced prior $y \sim \Pi$, Lemma 1 ensures the existence of at least one continuous stochastic process for each count process. All the proofs are reported in the Appendix.

LEMMA 1. *For every count stochastic process $y_0 \in \mathcal{C}$ satisfying Asssumption 1, there exists at least one continuous $y^* : \mathcal{S} \to \Re$ such that $y_0 = h(y^*)$.*

Defining an L^1 neighborhood around y_0 of size ϵ as

$$\eta_\epsilon(y_0) = \left\{ y : d_1(y_0, y) = \int |y_0(s) - y(s)| \mathrm{d}s < \epsilon \right\}, \tag{4}$$

we state the following Theorem on the prior support.

THEOREM 1. *Assuming the prior Π^* assigns positive probability to L^1 neighborhoods of any continuous function $y_0^* : \mathcal{S} \to \Re$, the prior Π induced through (3) assigns positive probability to L^1 neighborhoods of any $y_0 \in \mathcal{C}$ satisfying Assumption 1.*

In addition to showing large support of the prior, it is important to verify that the posterior distribution for y concentrates increasingly around the true process y_0 as the sample size increases. Theorem 2 provides sufficient conditions under which L^1 posterior consistency is obtained. Assumption 2 provides a space-filling regularity condition on the design.

Assumption 2. Let $\mathcal{S} = [0,1]^p$ and assume the n values of s_i arise from an in-fill design such that we can cover $\mathcal{S}$ with n L^∞ balls centered around $s_1, \ldots, s_n$ of size δ with $2\delta \in \left(n^{-1/p}, \lfloor n^{1/p} \rfloor^{-1}\right)$.

THEOREM 2. *Let $y \in \mathcal{C}$ be a count stochastic process with $y_i = y(s_i)$, for $i = 1, \ldots, n$ and $(s_1, \ldots, s_n)$ following Assumption 2. Letting $y_0 \in \mathcal{C}$ denote the true stochastic process and $y \sim \Pi$, then if $\Pi\{\eta_\epsilon(y_0)\} > 0$ for any ϵ and there exist sets $\{\mathcal{C}_n\}_{n=1}^\infty$ with $\mathcal{C}_n \in \mathcal{C}$ and $\mathcal{C}_n^C$ the complement of $\mathcal{C}_n$, where $\Pi\{\mathcal{C}_n^C\} < c_1 e^{-c_2 n}$, and c_1, c_2 positive constants, then*

$$\Pi\left\{\eta_\epsilon^C(y_0) \,|\, y_1, \ldots, y_n\right\} \to 0. \tag{5}$$

From Theorems 1 and 2, it follows that the prior proposed in equation (3) will lead to L^1 posterior consistency under Assumptions 1–2 as long as Π^* assigns positive probability to L^1 neighborhoods of any continuous function and negligible probability to $\mathcal{Y}_n^C = h^{-1}(\mathcal{C}_n^C)$ as n increases. Choi & Schervish (2007) showed that this condition holds, if $\mathcal{Y}_n^C$ has a particular form, for Π^* corresponding to orthogonal basis expansions or Gaussian processes with continuously

differentiable mean function and covariance having the form $k(s, s^{'}; \beta) = k_0(\beta|s - s^{'}|)$, where $s \in \Re$, $k_0(s)$ is a positive multiple of a density function four times continuously differentiable on $\Re$ and suitable hyperprior is chosen for β.

2·3. *Posterior computation*

We estimate the count process y at locations $s^{(N)}$ including observed locations $s^{(n)} = (s_1, \ldots, s_n)^T$ and additional locations of interest $s_{n+1}, \ldots, s_N$. Our rounded Gaussian process lets $y^* \sim \text{GP}(0, k)$, where $k(s, s') = \text{cov}\{y^*(s), y^*(s')\} = \tau_1 \exp(-\tau_2||s - s'||^2)$ is the covariance with $\tau_1^{-1} \sim \text{Ga}(a_{\tau_1}, b_{\tau_1})$ a scale parameter, $\tau_2^p \sim \text{Ga}(a_{\tau_2}, b_{\tau_2})$ and p the dimension of the domain $\mathcal{S}$. Here, τ_2^{-1} is a bandwidth parameter controlling smoothness, and this prior is motivated by the optimality results of van der Vaart & van Zanten (2009), though their theory does not apply directly in our setting. The resulting joint distribution of $y^{*(n)} = \{y^*(s_1), \ldots, y^*(s_n)\}^T$ is $N_n(0, \Sigma_n)$, with $\Sigma_n = \{\sigma_{ij}\}$ and $\sigma_{ij} = k(s_i, s_j)$.

Posterior computation can proceed via a Markov chain Monte Carlo algorithm.

Step 1 Sample $y^{*(n)}$ from $N_n(0, \Sigma_n)$ truncated to fall in a hyper-rectangle having $a_{y_i} \leq y^*(s_i) < a_{y_i+1}$.

Step 2 Sample τ_1^{-1} from conditional posterior distribution $\text{Ga}(a_{\tau_1} + n/2, b_{\tau_1} + y^{*\text{T}}\tau_1\Sigma_n y^*)$.

Step 3 Update τ_2 using a Metropolis-Hastings step.

Step 4 After burn-in, sample $y^*(s_{n+1}), \ldots, y^*(s_N)$ from the multivariate Gaussian conditional distribution.

In Step 1, Gibbs sampling can be used to update each $y^*(s_i)$ from its univariate truncated Gaussian conditional density, but this leads to slow mixing in our experience. Instead, we use the slice sampler of Liechty (2010), which samples multivariate Gaussian random variables restricted to a rectangular region. In step 3, the likelihood of $y^{(n)}$ marginalizing out $y^{*(n)}$ cannot be calculated analytically, so we rely on the multivariate normal likelihood of $y^{*(n)}$ in calculating the acceptance probability. It is well known that updating τ_2 conditionally on a latent Gaussian process can lead to stickiness, but due to the fact that our rounding approach minimizes differences between the observed y and the latent y^* we have not found this to be a major problem. Alternatively, one can improve mixing using the slice sampling approach of Murray & Adams (2010) with some additional complexity.

As for other Gaussian process models, we face a computational bottleneck and numerical instability as we evaluate y^* at increasing numbers of locations. Particularly when the process is observed at close locations and the covariance function favors smooth realizations, one obtains an ill-conditioned matrix, which can lead to large computational errors which degrade performance. There is a rich literature proposing solutions, with Banerjee et al. (2013) a recent example. A widely-used approximation represents the function as a linear combination of finitely many basis functions, leading to reduced instability problems and potentially improving Markov chain Monte Carlo mixing. Hence, along with the rounded Gaussian process, we implement an alternative that approximates y^* using penalized splines, with details on this approach provided in the Supplementary Material.

3. SIMULATION STUDY

A simulation study is conducted to assess the performance of the proposed approach, implemented using rounded Gaussian processes or P-splines, relative to several competitors. The first set of competitors initially treats the count measurements as continuous, assuming h to be

the identity function. The estimated continuous trajectory is then rounded in a second stage to produce an estimated count process. Such ad hoc two-stage approaches are simple to implement; we consider two-stage versions of rounded Gaussian processes and P-splines. A second approach treats the count measurements as ordered categorical using the Gaussian process ordinal regression model of Chu & Ghahramani (2005). This method faces complications when applied to counts and sparse ordered categorical data. In particular, letting $y(s_i) \in \{0, 1, \ldots, d\}$ for $i = 1, \ldots, n$ and $n_j = \sum_{i=1}^{n} 1_{\{y(s_i)=j\}}$, the total number of observations having value j, poor performance was obtained when any n_j was small, with lack of convergence when $n_j = 0$ for any $j \in \{0, 1, \ldots, d\}$. A third approach corresponds to Poisson regression with mean parameter $\lambda(s)$ estimated with a spline smoother as done by default by gam function of R library MASS. Lastly, we consider a simple interpolating step function defined as

$$f(s) = y_1 1_{s<s_2}(s) + \sum_{j=2}^{n} y_j 1_{s_j \le s < s_{j+1}}(s). \tag{6}$$

For our method, we considered the posterior median of $y(s)$. Simulations have been run under a wide variety of settings, leading to qualitatively similar results. We report the results for four scenarios. The first generates count stochastic processes from Poisson$\{\lambda(s)\}$, with $\lambda(s) = 2 + s/5 + \sin(s)$. In the second, y is generated by rounding a realization of a Gaussian process plus an error term,

$$y = h(y^*), \quad y^* \sim \text{GP}(\mu, k) + \epsilon, \tag{7}$$

with mean function $\mu(s) = 2 + \exp(s/5)$, covariance function $k(s, s^{'})$ squared exponential and $\epsilon(s)$ independent draws from $N(0, 2)$. These two cases do not satisfy Assumption 1, since infinitely many discontinuity points can occur. Under the third, we generate from a Poisson count process with rate parameter $1/2$ and in the fourth from (7) with $\epsilon = 0$.

For each case, we generated data on a equispaced grid of $1,000$ points between 0 and 20. Taking equispaced subsamples for different level of sparsity, namely of sizes $n = 25$, $n = 50$, $n = 100$, and $n = 500$, we estimate the trajectory on a fine grid for 500 replicates for each scenario and each method. Using Markov chain Monte Carlo simulation, we obtained draws from the posterior predictive distribution and used the median as our estimate. Methods are compared based on averaging the mean absolute deviation between the estimate and the true process across the replicates and grid points.

From Table 1, it is apparent that the proposed rounding approaches have the best overall performance. The Gaussian process ordinal regression model consistently has the worst performance. As expected the Poisson model with nonparametric mean performs well in scenario 1 but poorly in other cases, particularly when the sample size is not small. The interpolating step function has consistently poor performance except in scenario 3. The two-stage methods perform similarly to the proposed approaches in scenarios 1 and 2, but have substantially worse performance in scenarios 3 and 4. The two-stage methods have particularly poor performance when counts do not take a wide range of values, have values near zero, or tend to have many occurrences of the same value. In addition, the approach of rounding in a second stage can have unanticipated consequences in terms of inference on functionals, which may be unreliable and biased. Interestingly, the rounded P-splines approach has somewhat better performance than the rounded Gaussian process. Since rounded P-splines are also faster to implement, taking from 15 seconds for samples of size $n = 25$ to 30 seconds for samples of size $n = 500$ for 10,000 MCMC iterations in each of the simulated examples, we focus on this approach in the real data applications. We also compared the methods in terms of predictive mean absolute deviation, width and coverage of

Table 1. *Mean absolute deviation (and standard deviation) in simulation study of Section* 2·3

	Scenario 1				Scenario 2			
	$n = 25$	$n = 50$	$n = 100$	$n = 500$	$n = 25$	$n = 50$	$n = 100$	$n = 500$
RGP	2·10 (0·17)	2·04 (0·11)	1·93 (0·09)	1·02 (0·05)	2·27 (0·03)	2·12 (0·02)	1·98 (0·01)	0·93 (0·01)
GP	2·12 (0·02)	2·07 (0·02)	1·98 (0·01)	1·05 (0·01)	2·27 (0·03)	2·13 (0·02)	1·99 (0·01)	0·95 (0·01)
RPS	1·70 (0·09)	1·62 (0·07)	1·5 (0·05)	0·79 (0·03)	1·78 (0·11)	1·65 (0·06)	1·51 (0·05)	0·81 (0·03)
PS	1·70 (0·08)	1·63 (0·07)	1·51 (0·05)	0·8 (0·03)	1·81 (0·13)	1·69 (0·07)	1·55 (0·06)	0·83 (0·03)
GPOR	2·26 (0·33)	2·22 (0·26)	2·14 (0·21)	2·18 (0·14)	2·47 (0·31)	2·46 (0·31)	2·42 (0·24)	2·73 (0·14)
NPP	1·74 (0·08)	1·69 (0·06)	1·66 (0·05)	1·64 (0·04)	1·78 (0·1)	1·71 (0·06)	1·69 (0·06)	1·66 (0·05)
E	2·2 (0·18)	2·19 (0·13)	2·2 (0·1)	2·2 (0·08)	2·58 (0·2)	2·31 (0·13)	2·25 (0·11)	2·21 (0·06)
	Scenario 3				Scenario 4			
	$n = 25$	$n = 50$	$n = 100$	$n = 500$	$n = 25$	$n = 50$	$n = 100$	$n = 500$
RGP	0·12 (0·01)	0·07 (0·01)	0·05 (0·01)	0·01 (0·01)	0·34 (0·01)	0·25 (0·01)	0·17 (0·01)	0·05 (0·01)
GP	0·42 (0·01)	0·37(0·01)	0·34 (0·01)	0·21 (0·01)	0·58 (0·06)	0·52 (0·04)	0·47 (0·02)	0·26 (0·01)
RPS	0·14 (0·06)	0·08 (0·04)	0·05 (0·03)	0·02 (0·01)	0·28 (0·07)	0·19 (0·05)	0·13 (0·03)	0·05 (0·01)
PS	0·41 (0·08)	0·39 (0·06)	0·37 (0·06)	0·21 (0·03)	0·53 (0·06)	0·49 (0·04)	0·46 (0·03)	0·26 (0·01)
GPOR	2·88 (0·8)	3·03 (1·04)	3·26 (1·4)	3·65 (1·51)	2·25 (2·04)	2·6 (3·97)	4·74 (8·52)	5·9 (10·09)
NPP	0·27 (0·09)	0·26 (0·09)	0·26 (0·09)	0·26 (0·09)	0·56 (0·12)	0·56 (0·12)	0·56 (0·12)	0·56 (0·12)
E	0·18 (0·07)	0·09 (0·04)	0·05 (0·02)	0·01 (0)	1·11 (0·06)	0·59 (0·05)	0·31 (0·03)	0·09 (0·01)

RGP, rounded Gaussian process; GP, Gaussian process; RPS, rounded P-splines; PS, P-spline; GPOR, Gaussian process ordinal regression; NPP, nonparametric Poisson model; E, empirical interpolating step function.

predictive credible intervals and again observed better performance overall for the proposed approaches, with the competitors having high mean absolute deviation and poor coverage in at least one of the cases. Additional tables summarizing the results for predictive errors and predictive coverage are reported in the Supplementary Material.

4. REAL DATA APPLICATION

4·1. *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 (Behseta et al., 2005) to wavelet-based functional mixed models (Morris & Carroll, 2006).

Let $y_i(s)$ denote the count for subject i at time s, $y_{it} = y_i(s_{it})$, where s_{it} is the tth observation time for subject i, and $x_{it} = (x_{it1}, \ldots, x_{itp})^T$ predictors for subject i at the tth observation time. As a simple model motivated by the longitudinal tumor count and asthma inhaler use applications described below, we let

$$y_{it} = h(y_{it}^*), \quad y_{it}^* = \xi_i + b(s_{it}, x_{it})^T\theta + \epsilon_{it}, \quad \xi_i \sim Q, \quad \epsilon_{it} \sim N(0, \tau^{-1}), \tag{8}$$

where ξ_i is a subject-specific random effect, $b(\cdot)$ are basis functions that depend on time and predictors, θ are unknown basis coefficients, and ϵ_{it} is a residual which allows the counts to vary erratically from time to time about the smooth subject-specific mean curve. We use basis

expansions motivated by the success of rounded P-splines in our simulation. To allow the random effect distribution to be unknown, we choose a Dirichlet process prior (Ferguson, 1973), with $Q \sim \text{DP}(\alpha Q_0)$, with α a precision parameter and the base measure Q_0 chosen as $N(0, \psi)$ with $\psi \sim \text{Ga}(a_\psi, b_\psi)$. As commonly done we fix $\alpha = 1$. We additionally choose a hyperprior for the residual precision $p(\tau) \propto \tau^{-1}$ and for the basis coefficients $p(\theta)$, with the specific form of $p(\theta)$ depending on the context.

4·2. *Transgenic mouse bioassay application*

We first analyze data from a Tg.AC mouse bioassay study of pentaerythritol triacrylate, a chemical used in many industrial processes. Animals are randomized to a control or one of five dose groups each of size 30. The five dose groups are 0·75, 1·5, 3, 6, or 12 mg/kg. The number of skin papillomas on the back of each mouse is counted weekly for 26 weeks and it is of interest to compare the groups to see if there is an increase in tumorigenicity relative to control, while assessing dose response trend. Dunson & Herring (2005) analyzed these data through a Poisson-gamma frailty model. As motivated in §1·2, Poisson hierarchical models are quite restrictive and our focus here is on using the proposed model to improve robustness.

The only predictor for an animal is the dose group $x_i \in \{1, \ldots, G\}$ and we let $b(s_{it}, x_i)^T\theta = b(s_{it})^T\theta_{x_i}$ in expression (8) to allow a separate trajectory in time for each dose group, with $b(s)$ B-spline basis functions, θ_g basis coefficients specific to group g, and $p(\theta_g \mid \lambda) \propto \exp(-\lambda\theta_g^T P\theta_g/2)$ conditionally independent P-spline priors for each dose group. The prior is designed to only borrow information across dose groups in estimating smoothness parameter λ to avoid the possibility of having chemical effects in higher dose groups pull up the estimated tumor response in lower dose groups. To induce a heavy-tailed prior having appealing computational properties, we use a multilevel hierarchical prior for λ, with $\lambda \sim \text{Ga}(\nu/2, \delta\nu/2)$, $\delta \sim \text{Ga}(a_\delta, b_\delta)$ and $\psi \sim \text{Ga}(a_\psi, b_\psi)$. We do not expect to have substantial learning from the data about δ or ψ. Computational details are reported in the Supplementary Material.

As a global measure of toxicity, we use the average papilloma burden per group. The two lower dose groups showed no significant difference from the control group with the posterior mean of the average tumor burden $<$0·001 and the 95% credible intervals concentrated near zero. In the higher groups the average tumor burden grows with the dose level. Mean tumor burden and 95% credible intervals are 0·18 [0·06,0·39], 9·51 [9·21,9·80] and 12·33 [11·90,12·72] for the 3, 6 and 12 mg/kg dose group respectively. Cumulative tumor burdens along with the dose group-specific empirical means for each week are reported in Figure 2.

As a measure of time varying increase in papilloma burden, we computed the mean burden per dose group per week subtracting the average number for the control group. Posterior means and 95% credible bands are reported in Figure 3. The two lower dose groups are indistinguishable from control, with panel (a) of Figure 3 being a straght line equal to zero, while the 3, 6 and 12 mg/kg dose groups exhibit clear increases relative to control starting from the 17th, 9th and 8th week, respectively. Higher dosages lead to higher numbers of skin papillomas, and earlier onset of the first tumor. Our modeling approach allows us to estimate the average time of onset of first tumor, which occurs on the 27th, 14th and 11th week, for the three higher dose groups. In other groups, the typical mouse did not develop tumors prior to the end of the study.

Our overall conclusions agree with Dunson & Herring (2005). The group comparison results were also consistent with results from a frequentist generalized linear model analysis. We additionally implemented standard frequentist nonparametric tests for comparing groups based on summaries of the tumor trajectory data including time of first tumor and maximum tumor burden per animal. A p-value less than 0·001 for the Kruskal–Wallis rank sum test suggested strong evidence against equality among the dose groups in the maximum tumor burden per animal.

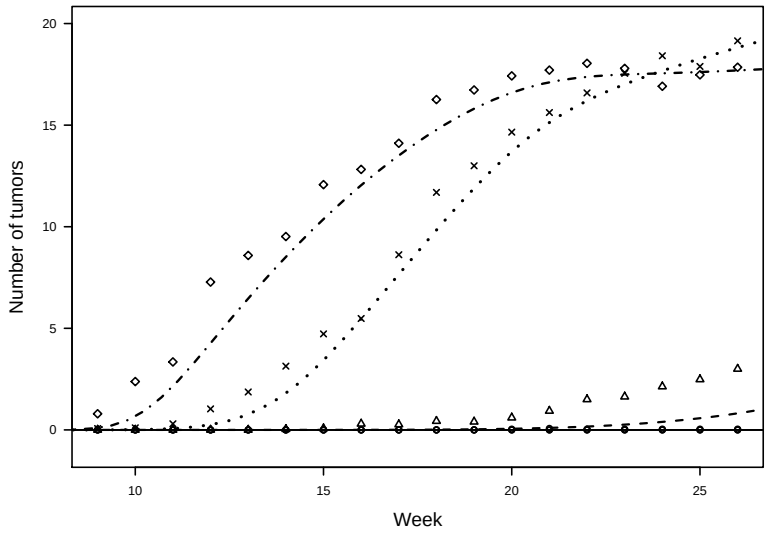

Fig. 2. Estimated cumulative mean tumor burden (lines) and weekly sample means (points) for the control group, 0·75 mg/kg and 1·5 mg/kg (solid line and circles), 3 mg/kg (dashed line and triangles), 6 mg/kg (dotted line and crosses) and 12 mg/kg (dash-dotted line and squares) dose groups.

Pairwise Wilcoxon tests were performed to test the equality of the maximum burden between each treated group and control, with one-sided alternatives of higher maximum burdens in the treated groups. P-values were less than 0·01 for the three higher groups while being 1 and 0·09 for the 0·75 and 1·5 mg/kg groups respectively. Similar results are obtained considering the time of development of the first tumor as a summary of the tumor trajectory. As partly illustrated in Figure 2, which shows the empirical and estimated mean tumor burdens in each group, the model has a good fit to the data.

4·3. *Asthma inhaler use application*

We additionally analyzed data on daily usage of albuterol asthma inhalers (Grunwald et al., 2011). Daily counts of inhaler use were recorded for a period between 36 and 122 days for 48 students previously diagnosed with asthma. The total number of observations was 5,209. As discussed by Grunwald et al. (2011), 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, in Fahrenheit degree/100, % humidity and barometric pressure, in mmHg/1000. We modify (8) to include these predictors in an additive model as follows.

$$y_{it} = h(y^*_{it}), \quad y^*_{it} = \xi_i + \sum_{j=1}^{4} b_j(x_{jt})^T \theta_j + \epsilon_{it}, \tag{9}$$

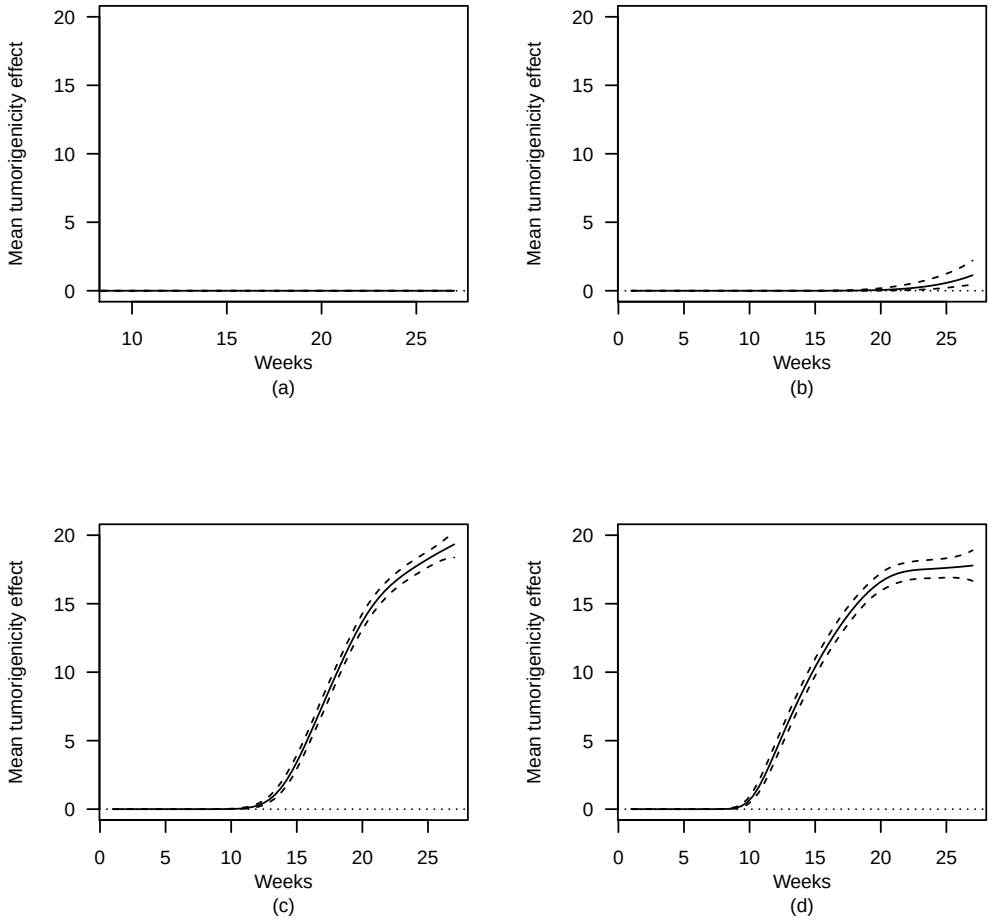

Fig. 3. Time varying chemical exposure posterior mean effect on tumorigenicity (continuous line) with 95% credible bands (dashed lines) for (a) 0·75 mg/kg and 1·5 mg/kg, (b) 3 mg/kg, (c) 6 mg/kg and (d) 12 mg/kg dose groups. Dotted line at zero corresponds to no effect of the chemical.

where ξ_i is a random effect modeled as in §4·2, b_j is a B-spline basis with θ_j the basis coefficients and $\epsilon_i \sim N(0, \tau^{-1}R)$, with R the correlation matrix arising from a first order autoregressive process with correlation parameter ρ. The prior for each θ_j is identical to the prior used for θ_g in §4·2 and 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 the Supplementary Material.

We ran our Markov chain Monte Carlo algorithm for 10,000 iterations with a 1,000 iteration burn-in discarded. Convergence and mixing were diagnosed by monitoring the non-linear effects of the different predictors at several values and also monitoring hyperparameters; The trace plots showed excellent mixing, with effective sample size over 9,000. Autocorrelation functions tend to drop near zero between lag 1 and 2. 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 $\xi_i = \mu_Q$, where μ_Q is the mean of the random effects distribution Q,

$$\begin{aligned}\mu_j(x_{jt}) &= E(y_{it} \,|\, x_{jt}, x_{j't} = 0, j' \neq j, \xi_i = \mu_Q, \theta, \tau, \rho) \\ &\approx \sum_{k=0}^{\lceil K \rceil} k[\Phi\{a_{k+1}; \mu_j^*(x_{jt}), \tau\} - \Phi\{a_k; \mu_j^*(x_{jt}), \tau\}], \end{aligned} \tag{10}$$

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

$$\mu_j^*(x_{jt}) = b_j(x_{jt})^{\mathrm{T}}\theta_j + \sum_{l \neq j} b_l(0)^{\mathrm{T}}\theta_l + \mu_Q, \tag{11}$$

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 4.

These data were previously analyzed by Grunwald et al. (2011) using a Faddy distribution with a log-linear mixed model for the mean,

$$\log E(y_{it} \,|\, x_t, \beta, u_i, e_{it}) = \sum_{j=1}^{p} x_{jt}\beta_j + u_i + e_{it}, \tag{12}$$

where u_i is a subject-specific random effect and e_{it} is a residual following a first-order autoregressive process. They estimated a coefficient of 0·013 for PM_{25}, which is close to zero with 95% confidence interval including zero. A Poisson log-linear model analysis yielded a similar coefficient of 0·014 but with a 50% wider confidence interval. Our approach, which is based on a substantially more flexible model that allows nonlinear effects and a nonparametric random effects distribution, produces results that are consistent with these earlier analyses.

5. DISCUSSION

We have proposed a simple new approach for modeling count stochastic processes based on rounding continuous stochastic processes. The general strategy is flexible and allows one to use 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 well-known limitations. We have explored some basic properties of rounding, but the primary contribution of this article is to introduce the idea that rounding is useful in this context. It is likely that some properties of the underlying continuous process, which are well known for Gaussian processes and for other standard cases, may carry over to the induced count process. However, this deserves further study. There are also interesting directions in terms of modeling counting processes corresponding to non-decreasing count processes via rounding non-decreasing continuous processes using monotone splines (Ramsay, 1998; Neelon & Dunson, 2004; Shively et al., 2009) and other constructions.

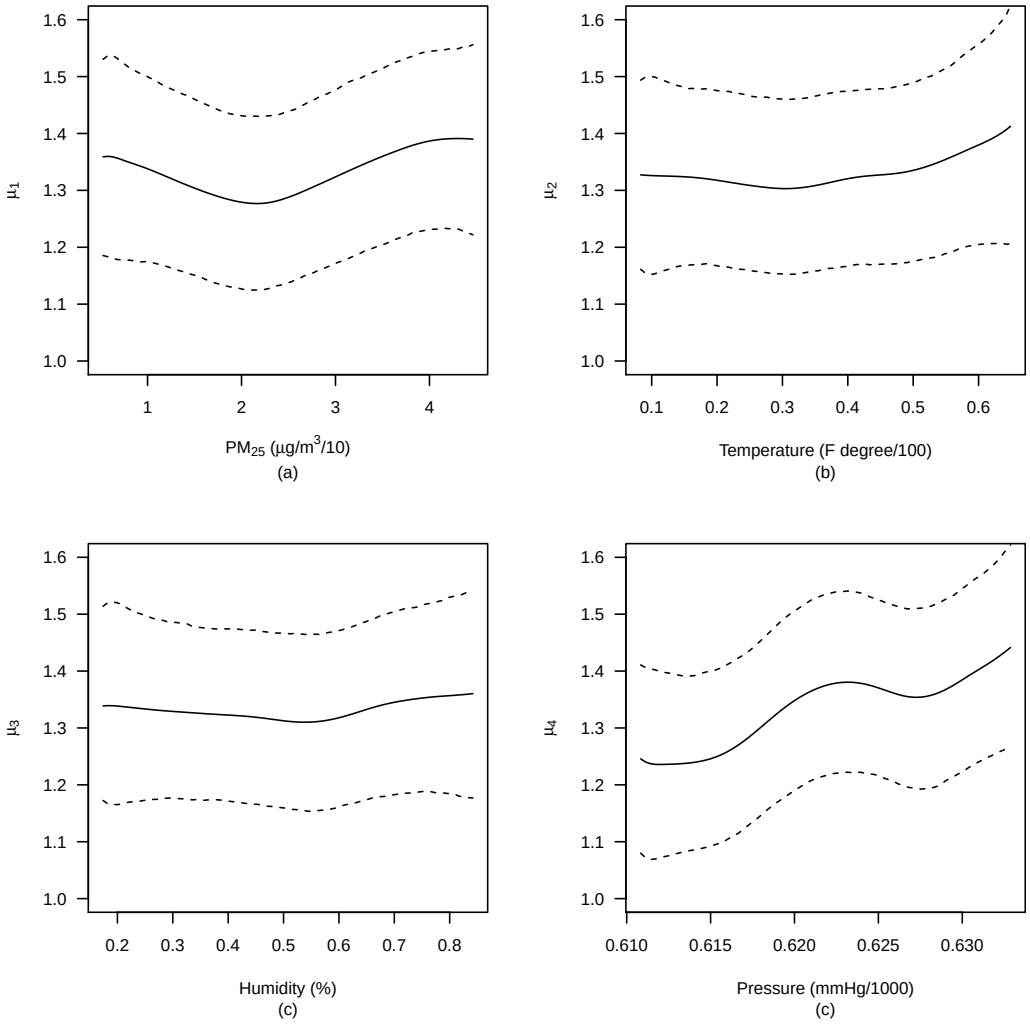

Fig. 4. Posterior mean and 95% pointwise credible bands for the effect of (a) concentration of PM_{25} pollutant, (b) average daily temperature, (c) % of humidity, and (d) barometric pressure on asthma inhaler use calculated with equation (10).

ACKNOWLEDGEMENTS

Comments and suggestions by the referees are gratefully acknowledged. Antonio Canale is also affiliated to the Collegio Carlo Alberto. This research was partially supported by grants from the National Institute of Environmental Health Sciences of the United States National Institutes of Health, the University of Padua, and Regione Piemonte.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes discussion on the covariance function of § 2·1, computational details for rounded P-splines of § 2·3, an additional table and

figure for simulation in § 3, and the computational details of the two applications discussed in § 4.

APPENDIX 1

Proof of Lemma 1. For any count stochastic process y_0 satisfying Assumption 1, we can partition the domain $\mathcal{S}$ into mutually disjoint sets $\mathcal{S}_l(y_0)$, with $y_0(s)$ constant within the interior of each $\mathcal{S}_l(y_0)$ and with unit increments at the boundaries. There are clearly infinitely many continuous functions $y^* : \mathcal{S} \to \Re$ satisfying the constraints (i) $y^*(s) \in [a_{y_0(s)}, a_{y_0(s)+1})$ for all $s \in \mathcal{S}$ and (ii) $y^*(s) = a_{y_0(s)}$ for $s \in \mathcal{B}(y_0)$. For all such y^*, we have $y_0 = h(y^*)$. □

Proof of Theorem 1. Theorem is an immediate consequence of Lemma 1 and of the following Lemma ensuring that the mapping h maintains L^1 neighborhoods.

LEMMA 2. *Suppose y^* and y_0^* are continuous and bounded by $M \in \Re$ such that $d_1(y^*, y_0^*) = \epsilon^*$, $y = h(y^*)$, $y_0 = h(y_0^*)$. Then, $y \in \eta_\epsilon(y_0)$ for all $\epsilon > \zeta(\epsilon^*; y_0^*)$, where $\zeta(\epsilon^*; y_0^*)$ is non-decreasing in ϵ^* having $\lim_{\epsilon^* \to 0} \zeta(\epsilon^*; y_0^*) = 0$.*

Proof of Lemma 2. Take $\mathcal{S} = [0,1]^p$ without loss of generality. Let $\{\mathcal{S}_l(y_0, y)\}_{l=1}^m$ the partition of $\mathcal{S}$ induced by $\{\mathcal{S}_l(y_0)\}_{l=1}^{m_0}$ and $\{\mathcal{S}_l(y)\}_{l=1}^{m_1}$ such that $y(s) = j_l$ and $y_0(s) = k_l$ for all $s \in \mathcal{S}_l(y_0, y)$ and some $j_l, k_l \in \mathcal{N}$. Let $\delta_l(y_0, y) = |j_l - k_l|$, for $l = 1, \ldots, m$ and $\lambda(\cdot)$ be the Lebesgue measure. Define

$$\zeta(\epsilon^*; y_0^*) = \sup_{y^* \in \eta_{\epsilon^*}(y_0^*)} \left\{ \max_{l=1,2,\ldots} [\delta_l\{y_0, h(y^*)\}] \sum_{l:\delta_l \neq 0} \lambda[\mathcal{S}_l\{y_0, h(y^*)\}] \right\}.$$

Clearly $y \in \eta_\epsilon(y_0)$ for all $\epsilon > \zeta(\epsilon^*; y_0^*)$ since

$$d_1(y_0, y) = \sum_{l=1}^m \delta_l(y_0, y)\lambda\{\mathcal{S}_l(y_0, y)\} \leq \zeta(\epsilon^*; y_0^*).$$

We show first that $\lim_{\epsilon^* \to 0} \zeta(\epsilon^*; y_0) = 0$. What follows holds for all $y^* \in \eta_{\epsilon^*}(y_0^*)$. Consider the general $y^* \in \eta_{\epsilon^*}(y_0^*)$. Since $\sum_{l:\delta_l \neq 0} \lambda[\mathcal{S}_l\{y_0, h(y^*)\}]$ is finite, $\zeta(\epsilon^*; y_0)$ goes to zero if $\max \delta_l\{y_0, h(y^*)\}$ goes to zero. Define $M_{\epsilon^*} = \max |y^*(s) - y_0^*(s)|$ and let $s_M = \arg\max |y^*(s) - y_0^*(s)|$ with s_M belonging to a given $\mathcal{S}_l$ where $y^*(s) \leq a_{j_l+1}$ and $y_0^*(s) \leq a_{k_l+1}$. For construction $|a_{l_l+1} - a_{k_l+1}| \leq M_{\epsilon^*}$ and so for $M_{\epsilon^*} \to 0$ we have $a_{l_l+1} = a_{k_l+1}$. Considering that $\max |y^*(s) - y_0^*(s)| \to 0$ then $|y^*(s) - y_0^*(s)| \to 0$ for all $s \in \mathcal{S}$ leading also to $\max \delta_l \to 0$. Whereas the absolute value of the difference $|y^*(s) - y_0^*(s)|$ is bounded and continuous we have that if $\int_{\mathcal{S}} |y^*(s) - y_0^*(s)| \, \mathrm{d}s$ goes to zero, also $\limsup_{\mathcal{S}} |y_0^*(s) - y^*(s)|$ goes to zero and hence also M_{ϵ^*}.

The fact that $\zeta(\cdot; y_0)$ is non decreasing follows directly from its definition. □

By Lemma 2 with suitable ϵ^* we have

$$\Pi\{\eta_\epsilon(y_0)\} = \Pi[h\{\eta_{\epsilon^*}(y_0^*)\}] = \Pi^*\{\eta_{\epsilon^*}(y_0^*)\} > 0.$$

Proof of Theorem 2. Since $y_0(s_i)$ is equal to the observed y_i for all i, we can rewrite the posterior (5) as

$$\begin{aligned}
\Pi\left\{y \in \eta_\epsilon^C(y_0) \mid y_1, \ldots, y_n\right\} &= \frac{\int_{\eta_\epsilon^C(y_0) \cap \mathcal{C}_n} \prod_{i=1}^n \delta_{y_i}(y_i) \mathrm{d}\Pi(y) + \int_{\eta_\epsilon^C(y_0) \cap \mathcal{C}_n^C} \prod_{i=1}^n \delta_{y_i}(y_i) \mathrm{d}\Pi(y)}{\int_{\mathcal{C}} \prod_{i=1}^n \delta_{y_i}(y_i) \mathrm{d}\Pi(y)} \\
&\leq \Phi_n + \frac{(1 - \Phi_n) \int_{\eta_\epsilon^C(y_0) \cap \mathcal{C}_n} \prod_{i=1}^n \delta_{y_i}(y_i) \mathrm{d}\Pi(y) + \int_{\eta_\epsilon^C(y_0) \cap \mathcal{C}_n^C} \prod_{i=1}^n \delta_{y_i}(y_i) \mathrm{d}\Pi(y)}{\int_{\mathcal{C}} \prod_{i=1}^n \delta_{y_i}(y_i) \mathrm{d}\Pi(y)} \\
&= \Phi_n + \frac{I_{1,n}(y_1, \ldots, y_n) + I_{2,n}(y_1, \ldots, y_n)}{I_{3,n}(y_1, \ldots, y_n)},
\end{aligned}$$

where δ_a is a delta mass at a, Φ_n is a test function and $\mathcal{C}_n$ is a sieve that grows eventually to the whole space $\mathcal{C}$. It suffices to show that

$$\Phi_n \to 0, \tag{A1}$$
$$e^{\beta_1 n} I_{1,n}(y_1, \ldots, y_n) \to 0, \tag{A2}$$
$$e^{\beta_2 n} I_{2,n}(y_1, \ldots, y_n) \to 0, \tag{A3}$$
$$e^{\beta n} I_{3,n}(y_1, \ldots, y_n) \to \infty \tag{A4}$$

with $\beta < \min\{\beta_1, \beta_2\}$.

Denote $\lfloor a \rfloor$ the integer part of a and let $\mathcal{S} = \bigcup_{j=1}^{\lfloor n^{1/p} \rfloor^p} \mathcal{G}_j$ with $\mathcal{G}_j$ an L^∞ ball of size $0.5(\lfloor n^{1/p} \rfloor)^{-1}$ and center s'_j, where the centers are chosen on a grid so that $\lfloor n^{1/p} \rfloor^p$ balls cover $\mathcal{S}$ and each $\mathcal{G}_j$ contains at least one element of $(s_1, \ldots, s_n)^T$ under Assumption 2. Define $X_i = 1\{y(s_i) = y_0(s'_j)\}$ with s'_j being the centroid of the $\mathcal{G}_j$ in which s_i is contained. Let $\Phi_n = 1\{\sum_{i=1}^n X_i < n\}$ the test on the set

$$\mathcal{C}_n = \left\{ y : y \text{ is constant in } \mathcal{G}_j, \text{for all } j = 1, \ldots, \lfloor n^{1/p} \rfloor^p, ||y||_\infty < M_n \right\} \tag{A5}$$

with $M_n = \mathcal{O}(n^\alpha)$ and $1/2 < \alpha < 1$. The first condition on the sieve governs the regularity of the process while the second gives an upper bound for the infinity norm as in Choi & Schervish (2007). The true y_0 belongs to $\mathcal{C}_n$ for a given n and hence for n sufficiently large the test functions have exactly zero type I and type II probability. From this (A1) is directly verified. We continue to prove (A2). By Fubini's theorem we have

$$\begin{aligned} E_{y_0}\{I_{1,n}(y_1, \ldots, y_n)\} &= E_{y_0}\left\{ (1 - \Phi_n) \int_{\eta_\epsilon^C(y_0) \cap \mathcal{C}_n^C} \prod_{i=1}^n \delta_{y_i}(y_i) \mathrm{d}\Pi(y) \right\} \\ &= \int_{\eta_\epsilon^C(y_0) \cap \mathcal{C}_n^C} E_y\{(1 - \Phi_n)\} = 0 \end{aligned}$$

where the final equality is directly verified by the test construction. Next we prove (A3). Again by Fubini's theorem we have

$$E_{y_0}\{I_{2,n}(y_1, \ldots, y_n)\} = E_{y_0}\left\{ \int_{\eta_\epsilon^C(y_0) \cap \mathcal{C}_n^C} \prod_{i=1}^n \delta_{y_i}(y_i) \mathrm{d}\Pi(y) \right\} \leq \Pi(\mathcal{C}_n^C) \leq c_1 e^{-c_2 n},$$

and hence for $\beta_2 < c_2$, $e^{\beta_2 n} I_{2,n}(y_1, \ldots, y_n) \to 0$. Finally the prior positivity of Π makes $I_{3,n}(y_1, \ldots, y_n)$ to be positive. This proves also (A4) and concludes the proof. □

REFERENCES

BANERJEE, A., DUNSON, D. B. & TOKDAR, S. T. (2013). Efficient gaussian process regression for large datasets. *Biometrika* **100**, 75–89.

BANERJEE, S., GELFAND, A. E., FINLEY, A. O. & SANG, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society, Series B: Statistical Methodology* **70**, 825–848.

BEHSETA, S., KASS, R. E. & WALLSTROM, G. L. (2005). Hierarchical models for assessing variability among functions. *Biometrika* **92**, 419–434.

CANALE, A. & DUNSON, D. B. (2011). Bayesian kernel mixtures for counts. *Journal of the American Statistical Association* **106**, 1528–1539.

CHOI, T. & SCHERVISH, M. J. (2007). On posterior consistency in nonparametric regression problems. *Journal of Multivariate Analysis* **98**, 1969–1987.

CHU, W. & GHAHRAMANI, Z. (2005). Gaussian process for ordinal regression. *Journal of Machine learning Research* **6**, 1019–1041.

DUNSON, D. B. & HERRING, A. H. (2005). Bayesian latent variable models for mixed discrete outcomes. *Biostatistics* **6**, 11–25.

ERHARD, V. & CZADO, C. (2009). Sampling count variables with specified Pearson correlation - a comparison between a naive and a C-vine sampling approach. In *Dependence Modeling - Handbook on Vine Copulae*, D. Kurowicka & H. Joe, eds. World Scientific, pp. 73–87.

FADDY, M. J. (1997). Extended Poisson process modeling and analysis of count data. *Biometrical Journal* **39**, 431–440.

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1**, 209–230.

FRÜHWIRTH-SCHNATTER, S. & WAGNER, H. (2006). Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modeling. *Biometrika* **93**, 827–841.

GENEST, C. & NESLEHOVA, J. (2007). A primer on copulas for count data. *Astin Bulletin* **37**, 475–515.

GHOSAL, S. & ROY, A. (2006). Posterior consistency of Gaussian process prior for nonparametric binary regression. *The Annals of Statistics* **34**, 2413–2429.

GRUNWALD, G. K., BRUCE, S. L., JIANG, L., STRAND, M. & RABINOVITCH, N. (2011). A statistical model for under- or overdispersed clustered and longitudinal count data. *Biometrical Journal* **53**, 578–594.

KACHOUR, M. & YAO, J. F. (2009). First order rounded integer-valued autoregressive (RINAR(1)) process. *Journal of time series analysis* **30**, 417–448.

LIECHTY, M. W. (2010). Multivariate normal slice sampling. *Journal of Computational and Graphical Statistics* **19**, 281–294.

MORRIS, J. S. & CARROLL, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society, Series B: Statistical Methodology* **68**, 179–199.

MURRAY, I. & ADAMS, R. P. (2010). Slice sampling covariance hyperparameters of latent Gaussian models. In *Advances in Neural Information Processing Systems 23*, J. Lafferty, C. K. I. Williams, R. Zemel, J. Shawe-Taylor & A. Culotta, eds.

NEELON, B. & DUNSON, D. B. (2004). Bayesian isotonic regression and trend analysis. *Biometrics* **60**, 398–406.

NIKOLOULOPOULOS, A. & KARLIS, D. (2010). Regression in a copula model for bivariate count data. *Journal of Applied Statistics* **37**, 1555–1568.

RAMSAY, J. O. (1998). Estimating smooth monotone functions. *Journal of the Royal Statistical Society, Series B: Statistical Methodology* **60**, 365–375.

RODRÍGUEZ, A., DUNSON, D. B. & GELFAND, A. E. (2010). Latent stick-breaking processes. *Journal of the American Statistical Association* **105**, 647–659.

RUE, H., MARTINO, S. & CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B: Statistical Methodology* **71**, 319–392.

SHIVELY, T. S., SAGER, T. W. & WALKER, S. G. (2009). A Bayesian approach to non-parametric monotone function estimation. *Journal of the Royal Statistical Society, Series B: Statistical Methodology* **71**, 159–175.

SHMUELI, G., MINKA, T. P., KADANE, J. B., BORLE, S. & BOATWRIGHT, P. (2005). A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution. *Journal Of The Royal Statistical Society Series C* **54**, 127–142.

VAN DER VAART, A. W. & VAN ZANTEN, J. H. (2009). Adaptive Bayesian estimation using a Gaussian random field with inverse gamma bandwidth. *The Annals of Statistics* **37**, 2655–2675.

WILSON, A. & GHAHRAMANI, Z. (2010). Copula processes. In *Advances in Neural Information Processing Systems 23*, J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. Zemel & A. Culotta, eds.