ISSN 2279-9362

Collegio Carlo Alberto

Nonparametric Bayes modelling of count processes

Antonio Canale
David B. Dunson

No. 318
December 2013

Carlo Alberto Notebooks

www.carloalberto.org/research/working-papers

Biometrika (2013), *pp.* 1–16
Printed in Great Britain
doi: 10.1093/biomet/ast037

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

AND DAVID B. DUNSON

Department of Statistical Science, Duke University, Box 90251, Durham, North Carolina 27708, U.S.A.
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 applied contexts, and hierarchical Poisson models make restrictive assumptions about overdispersion in marginal distributions. In this article we propose a class of nonparametric Bayes count process models, constructed through rounding real-valued underlying processes. The proposed class of models accommodates situations in which separate count-valued functional data are observed for each subject under study. Theoretical results on large support and posterior consistency are established, and computational algorithms are developed based on Markov chain Monte Carlo simulation. The methods are evaluated via simulation and illustrated by application to longitudinal tumour counts and to asthma inhaler usage.

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

1. INTRODUCTION

A stochastic process $y = \{y(s) : s \in \mathcal{S}\}$ is a collection of random variables indexed by $s \in \mathcal{S}$, where the domain $\mathcal{S}$ usually corresponds to a set of times or spatial locations and $y(s)$ is 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 encompassing Gaussian processes and Lévy processes, such as the Poisson, Wiener, beta or gamma processes. Gaussian processes provide a convenient and well-studied choice when $y : \mathcal{S} \to \mathbb{R}$ 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) developed improved methods for posterior computation, while Ghosal & Roy (2006) and van der Vaart & van Zanten (2009) studied asymptotic properties, including posterior consistency and rates of convergence. The Gaussian process is appealing in that it provides a prior which can be specified to generate functions that are within an arbitrarily small neighbourhood

of any continuous function with positive probability (Ghosal & Roy, 2006), while also being computationally convenient.

Our interest is focused on the case where $y : \mathcal{S} \to \mathcal{N} = \{0, \ldots, \infty\}$, so that y is a count-valued stochastic process over the domain $\mathcal{S}$. There are many applications of such processes, for example epidemiological studies monitoring a count biomarker or health response of patients over time and ecological studies recording the number of birds of a given species observed at different locations. Although there is a rich literature on models for longitudinal and spatial data, most models rely on Poisson hierarchical specifications. For example, Frühwirth-Schnatter & Wagner (2006) consider $y(s) \sim \text{Po}\{\lambda(s)\}$, where the Poisson mean $\lambda(s)$ varies 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 following 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 that it limits the variance to equal the mean, and overdispersion is introduced in marginalizing out the latent processes. This leads to a confounding of the dependence structure with the degree of overdispersion in the marginals, as both are induced through the latent process. Such modelling frameworks cannot accommodate correlated count data that are underdispersed, and substantial bias can result for non-Poisson overdispersed data. Relying on a hierarchical Faddy model (Faddy, 1997), Grunwald et al. (2011) developed methods that generalize the Poisson distribution to accommodate under- and overdispersed 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 underdispersion. 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 underdispersion 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 that are directly applicable to count stochastic processes have not yet been developed. 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 modelled via a stick-breaking prior. They considered a spatial count process application, with the marginal modelled 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. In this article, we instead propose a class

of stochastic processes that map a real-valued stochastic process $y^* : \mathcal{S} \to \mathbb{R}$ 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 $\mathcal{S} \subset \mathbb{R}^p$ compact and $\mathcal{C}$ being the set of all $\mathcal{S} \to \mathcal{N}$ functions that satisfy 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 $y(s)$ being constant in the interior of each set $\mathcal{S}_l(y)$ and having 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 can easily be relaxed.

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

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

as well as the joint probability mass functions

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

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

In introducing the Dirichlet process, Ferguson (1973) mentioned three appealing characteristics of nonparametric Bayes priors: 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 the large support characteristic, we would like to choose a Π that allocates positive probability to arbitrarily small neighbourhoods 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 allowing one to assume that y belongs to a prespecified subset of $\mathcal{C}$ with probability 1, 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 neighbourhoods of any true $y_0 \in \mathcal{C}$.

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

$$y = h(y^*), \quad y^* \sim \Pi^*, \tag{1}$$

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

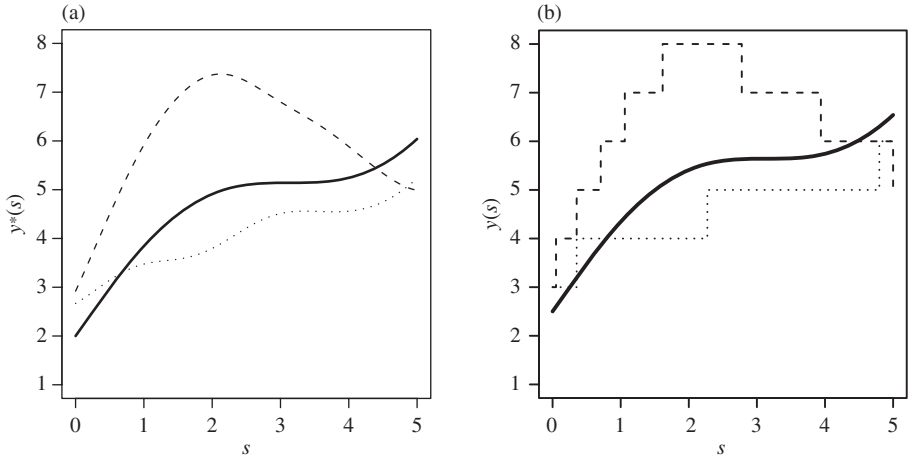

Fig. 1. Effect of the thresholding operator h in (1): (a) samples from a Gaussian process (dashed and dotted lines) with mean function $\mu(s) = 2 + \sin(s) + s$ (solid line) and squared exponential covariance function; (b) rounded versions of the realizations in panel (a) (dashed and dotted lines) together with the induced mean function (solid line).

restricting y to lie in $\mathcal{C}$. Hence we focus on a rounding operator that sets $y(s) = 0$ if $y^*(s) < 0$ and $y(s) = j$ if $j - 1 \leqslant y^*(s) < j$ for $j = 1, \ldots, \infty$. In other words, negative values will be mapped to zero, which is the closest nonnegative integer, while positive values will be rounded up to the nearest integer. This type of restricted rounding ensures that $y(s)$ will be a nonnegative integer. Using a fixed rounding function h in (1), 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 $y^*(s) \in A_j = [a_j, a_{j+1})$, where $a_0 < \cdots < a_\infty$, and we focus on the case with $a_0 = -\infty$ and $a_j = j - 1$ for $j = 1, \ldots, \infty$.

This construction is particularly suitable for modelling 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 subregions 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 the longitudinal tumour count study of § 4·2.

Figure 1 illustrates the prior through showing, in panel (a), realizations of the underlying stochastic process and, in panel (b), the resulting count process after applying the rounding operator. The thick solid lines represent the mean functions of the real-valued process and 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 $\text{cov}\{y^*(s), y^*(s)\}$. In the Supplementary Material we present some plots comparing the covariance of the original process with that of the induced process.

In certain applications, count data can naturally be viewed as arising through integer-valued rounding of an underlying continuous process. For example, in the longitudinal tumour count study described in § 4·2, as distinguishing individual tumours tends to be difficult, it is natural to posit a continuous time-varying tumour burden, with tumours fusing together and falling off over time. Moreover, while tumour biologists attempt to make an accurate count when collecting data, measurement errors are unavoidable, and one can naturally take this into account by assuming a smoothly varying continuous tumour burden specific to each animal, with measurement errors and rounding producing the observed tumour counts. However, even when the existence of an underlying continuous process does not have a clear motivation in terms of the applied context, our proposed formulation still leads to a highly flexible and computationally convenient model.

2·2. *Properties*

The mapping function $h(\cdot)$ in (1) 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. The proofs of all results in this section are given in the Appendix.

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

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

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

we have the following theorem on the prior support.

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

In addition to showing the large support property 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 holds. Assumption 2 provides a space-filling regularity condition on the design.

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

THEOREM 2. *Let $y \in \mathcal{C}$ be a count stochastic process with $y_i = y(s_i)$ for $i = 1, \ldots, n$, where $s_1, \ldots, s_n$ are as in Assumption 2. Let $y_0 \in \mathcal{C}$ denote the true stochastic process and let $y \sim \Pi$. Suppose that $\Pi\{\eta_\epsilon(y_0)\} > 0$ for any ϵ and there exist sets $\{\mathcal{C}_n\}_{n=1}^\infty$ with $\mathcal{C}_n \in \mathcal{C}$ and $\Pi\{\mathcal{C}_n^{\mathrm{C}}\} <$*

$c_1 \exp(-c_2 n)$, where $\mathcal{C}_n^{\mathrm{C}}$ denotes the complement of $\mathcal{C}_n$ and c_1, c_2 are positive constants. Then

$$\Pi\left\{\eta_\epsilon^{\mathrm{C}}(y_0)\,\middle|\, y_1, \ldots, y_n\right\} \to 0. \tag{2}$$

From Theorems 1 and 2, it follows that the prior proposed in (1) will lead to L^1 posterior consistency under Assumptions 1 and 2 as long as Π^* assigns positive probability to L^1 neighbourhoods of any continuous function and negligible probability to $\mathcal{Y}_n^{\mathrm{C}} = h^{-1}(\mathcal{C}_n^{\mathrm{C}})$ as n increases. Choi & Schervish (2007) showed that this condition holds if $\mathcal{Y}_n^{\mathrm{C}}$ 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 $s \in \mathbb{R}$, $k_0(s)$ is a positive multiple of a density function that is four times continuously differentiable on $\mathbb{R}$, and a suitable hyperprior is chosen for β.

2·3. *Posterior computation*

We estimate the count process y at locations $s^{(N)}$, including observed locations $s^{(n)} = (s_1, \ldots, s_n)^{\mathrm{T}}$ and additional locations of interest $s_{n+1}, \ldots, s_N$. Our rounded Gaussian process has $y^* \sim \mathrm{GP}(0, k)$ with $k(s, s') = \mathrm{cov}\{y^*(s), y^*(s')\} = \tau_1 \exp(-\tau_2 \|s - s'\|^2)$, where $\tau_1^{-1} \sim \mathrm{Ga}(a_{\tau_1}, b_{\tau_1})$ is a scale parameter and $\tau_2^p \sim \mathrm{Ga}(a_{\tau_2}, b_{\tau_2})$, with p being 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), although their theory does not apply directly in our setting. The resulting joint distribution of $y^{*(n)} = \{y^*(s_1), \ldots, y^*(s_n)\}^{\mathrm{T}}$ 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 with $a_{y_i} \leqslant y^*(s_i) < a_{y_i+1}$.

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

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 in our experience this leads to slow mixing. 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 owing 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 by using the slice-sampling approach of Murray & Adams (2010) with some additional complexity.

As with 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 favours smooth realizations, one obtains an ill-conditioned matrix, which can lead to large computational errors that degrade performance. There is a rich literature proposing solutions to this problem, with the paper of Banerjee et al.

(2013) being a recent contribution. In a widely used approximation, the function is represented as a linear combination of finitely many basis functions, leading to reduced instability and potentially improved Markov chain Monte Carlo mixing. Hence, along with the rounded Gaussian process, we implement an alternative that approximates y^* using penalized splines. Details of this approach are provided in the Supplementary Material.

3. Simulation study

A simulation study is conducted to assess the performance of the proposed approach. The approach is implemented using rounded Gaussian processes or P-splines, and it is compared with 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 variates, 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, taking $y(s_i) \in \{0, 1, \ldots, d\}$ for $i = 1, \ldots, n$ and defining $n_j = \sum_{i=1}^{n} 1_{\{y(s_i)=j\}}$, 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 the mean parameter $\lambda(s)$ estimated with a spline smoother, as done by default by the gam function of the R library MASS. Lastly, we consider a simple interpolating step function defined as

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

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. Scenario 1 generates count stochastic processes from $\text{Po}\{\lambda(s)\}$, with $\lambda(s) = 2 + s/5 + \sin(s)$. In scenario 2, y is generated by rounding a realization of a Gaussian process plus an error term,

$$y = h(y^*), \quad y^* \sim \text{GP}(\mu, k) + \epsilon, \tag{3}$$

with mean function $\mu(s) = 2 + \exp(s/5)$, squared exponential covariance function $k(s, s')$ and $\epsilon(s)$ consisting of independent draws from $N(0, 2)$. These two cases do not satisfy Assumption 1, since infinitely many discontinuity points can occur. In scenario 3 we generate from a Poisson count process with rate parameter $1/2$, and in scenario 4 we generate from (3) with $\epsilon = 0$.

In each case, we generated data on a equispaced grid of 1000 points between 0 and 20. Taking equispaced subsamples for different levels of sparsity, of sizes $n = 25$, 50, 100 and 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.

Table 1 shows that the proposed rounding approaches have the best overall performance. The Gaussian process ordinal regression model consistently performed worst. As expected, the Poisson model with nonparametric mean performs well in scenario 1 but poorly in the other cases, particularly when the sample size is not small. The interpolating step function showed consistently poor performance except in scenario 3. The two-stage methods perform similarly to the proposed

Table 1. *Mean absolute deviation and standard deviation (in parentheses) in the simulation study of § 3*

		Scenario 1		
	$n = 25$	$n = 50$	$n = 100$	$n = 500$
RGP	2·10 (0·17)	2·04 (0·11)	1·93 (0·09)	1·02 (0·05)
GP	2·12 (0·02)	2·07 (0·02)	1·98 (0·01)	1·05 (0·01)
RPS	1·70 (0·09)	1·62 (0·07)	1·50 (0·05)	0·79 (0·03)
PS	1·70 (0·08)	1·63 (0·07)	1·51 (0·05)	0·80 (0·03)
GPOR	2·26 (0·33)	2·22 (0·26)	2·14 (0·21)	2·18 (0·14)
NPP	1·74 (0·08)	1·69 (0·06)	1·66 (0·05)	1·64 (0·04)
E	2·20 (0·18)	2·19 (0·13)	2·20 (0·1)	2·20 (0·08)
		Scenario 2		
	$n = 25$	$n = 50$	$n = 100$	$n = 500$
RGP	2·27 (0·03)	2·12 (0·02)	1·98 (0·01)	0·93 (0·01)
GP	2·27 (0·03)	2·13 (0·02)	1·99 (0·01)	0·95 (0·01)
RPS	1·78 (0·11)	1·65 (0·06)	1·51 (0·05)	0·81 (0·03)
PS	1·81 (0·13)	1·69 (0·07)	1·55 (0·06)	0·83 (0·03)
GPOR	2·47 (0·31)	2·46 (0·31)	2·42 (0·24)	2·73 (0·14)
NPP	1·78 (0·1)	1·71 (0·06)	1·69 (0·06)	1·66 (0·05)
E	2·58 (0·2)	2·31 (0·13)	2·25 (0·11)	2·21 (0·06)
		Scenario 3		
	$n = 25$	$n = 50$	$n = 100$	$n = 500$
RGP	0·12 (0·01)	0·07 (0·01)	0·05 (0·01)	0·01 (0·01)
GP	0·42 (0·01)	0·37(0·01)	0·34 (0·01)	0·21 (0·01)
RPS	0·14 (0·06)	0·08 (0·04)	0·05 (0·03)	0·02 (0·01)
PS	0·41 (0·08)	0·39 (0·06)	0·37 (0·06)	0·21 (0·03)
GPOR	2·88 (0·8)	3·03 (1·04)	3·26 (1·40)	3·65 (1·51)
NPP	0·27 (0·09)	0·26 (0·09)	0·26 (0·09)	0·26 (0·09)
E	0·18 (0·07)	0·09 (0·04)	0·05 (0·02)	0·01 (0)
		Scenario 4		
	$n = 25$	$n = 50$	$n = 100$	$n = 500$
RGP	0·34 (0·01)	0·25 (0·01)	0·17 (0·01)	0·05 (0·01)
GP	0·58 (0·06)	0·52 (0·04)	0·47 (0·02)	0·26 (0·01)
RPS	0·28 (0·07)	0·19 (0·05)	0·13 (0·03)	0·05 (0·01)
PS	0·53 (0·06)	0·49 (0·04)	0·46 (0·03)	0·26 (0·01)
GPOR	2·25 (2·04)	2·60 (3·97)	4·74 (8·52)	5·90 (10·09)
NPP	0·56 (0·12)	0·56 (0·12)	0·56 (0·12)	0·56 (0·12)
E	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-spline; PS, P-spline; GPOR, Gaussian process ordinal regression; NPP, nonparametric Poisson model; E, empirical interpolating step function.

approaches in scenarios 1 and 2, but have substantially worse performance in scenarios 3 and 4. The two-stage methods show particularly poor performance when the 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 Markov chain Monte Carlo 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, as well as width and coverage of predictive credible intervals, and again observed better overall performance of the proposed

approaches, with the competitors having high mean absolute deviation and poor coverage in at least one of the scenarios. Additional tables summarizing the results for predictive errors and predictive coverage are presented in the Supplementary Material.

4. Applications

4·1. *Count functional data*

We have concentrated on the case where a single count process y is observed at locations $s = (s_1, \ldots, s_n)^{\mathrm{T}}$. In many applications, however, there are multiple related count processes $\{y_i : i = 1, \ldots, n\}$, with the ith process observed at locations $s_i = (s_{i1}, \ldots, s_{in_i})^{\mathrm{T}}$. 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, let $y_{it} = y_i(s_{it})$ where s_{it} is the tth observation time for subject i, and let $x_{it} = (x_{it1}, \ldots, x_{itp})^{\mathrm{T}}$ be predictors for subject i at the tth observation time. As a simple model motivated by the longitudinal tumour count and asthma inhaler use applications described below, we let

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

where ξ_i is a subject-specific random effect, the $b(\cdot)$ are basis functions that depend on time and predictors, θ is a vector of 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), $Q \sim \mathrm{DP}(\alpha Q_0)$, where α is a precision parameter and the base measure Q_0 is chosen as $N(0, \psi)$ with $\psi \sim \mathrm{Ga}(a_\psi, b_\psi)$. As is commonly done, we fix $\alpha = 1$. Additionally, we 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*

We first analyse data from a Tg.AC mouse bioassay study of pentaerythritol triacrylate, a chemical used in many industrial processes. Animals were randomized to a control group or to 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 was counted weekly for 26 weeks, and it is of interest to compare the groups to see if there is an increase in tumourigenicity relative to control, while assessing dose-response trend. Dunson & Herring (2005) analysed these data through a Poisson-gamma frailty model. As remarked earlier, Poisson hierarchical models are quite restrictive, and our aim here is to use 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 $b(s_{it}, x_i)^{\mathrm{T}}\theta = b(s_{it})^{\mathrm{T}}\theta_{x_i}$ in expression (4) to allow a separate trajectory in time for each dose group. Here the $b(s)$ are B-spline basis functions, the θ_g are basis coefficients specific to group g, and $p(\theta_g \mid \lambda) \propto \exp(-\lambda\theta_g^{\mathrm{T}} P \theta_g/2)$ are conditionally independent P-spline priors for each dose group. The prior is designed to only borrow information across dose groups in estimating the smoothness

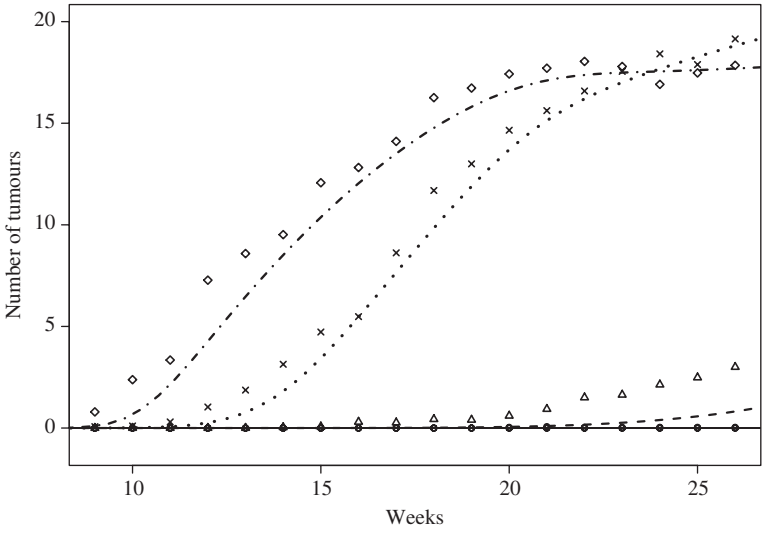

Fig. 2. Estimated cumulative mean tumour burden (lines) and weekly sample means (points) for: the control, 0·75 mg/kg dose and 1·5 mg/kg dose groups (solid line and circles); the 3 mg/kg dose group (dashed line and triangles); the 6 mg/kg dose group (dotted line and crosses); and the 12 mg/kg dose group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

parameter λ, to avoid the possibility of having chemical effects in higher dose groups pull up the estimated tumour response in lower dose groups. To induce a heavy-tailed prior with 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. The 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 tumour burden being less than 0·001 and the 95% credible intervals concentrated near zero. For the higher dose groups, the average tumour burden grows with the dose level. Mean tumour 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 groups, respectively. Cumulative tumour burdens along with the dose group-specific empirical means for each week are plotted in Fig. 2.

As a measure of the time-varying increase in papilloma burden, we took the mean burden per dose group per week and subtracted the average number for the control group. Posterior means and 95% credible bands are reported in Fig. 3. The two lower dose groups are indistinguishable from the control, so that Fig. 3(a) shows a horizontal line at zero, while the 3, 6 and 12 mg/kg dose groups exhibit clear increases relative to control starting from the 17th, 9th and 8th weeks, respectively. Higher doses of the chemical are associated with higher numbers of skin papillomas and earlier onset of the first tumour. Our modelling approach allows us to estimate the average time of onset of the first tumour, which is in the 27th, 14th and 11th weeks for the three higher dose groups. For the other groups, the typical mouse did not develop tumours before the end of the study.

Our overall conclusions agree with those of Dunson & Herring (2005). The group comparison results are also consistent with results obtained from a frequentist generalized linear model analysis. We additionally implemented standard frequentist nonparametric tests for comparing groups based on summaries of the tumour trajectory data, including time of first tumour and maximum tumour burden per animal. A p-value of less than 0·001 for the Kruskal–Wallis rank

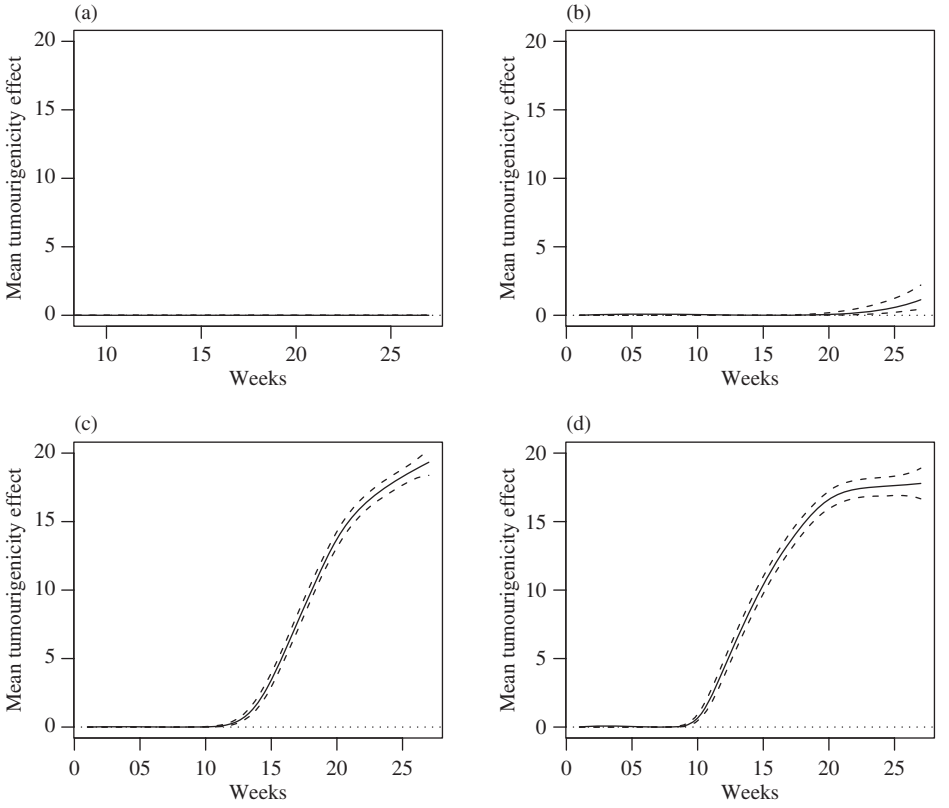

Fig. 3. Time-varying chemical exposure posterior mean effect on tumourigenicity (solid line) with 95% credible bands (dashed lines) for: (a) the 0·75 mg/kg and 1·5 mg/kg dose groups; (b) the 3 mg/kg dose group; (c) the 6 mg/kg dose group; and (d) the 12 mg/kg dose group. The dotted horizontal line at zero corresponds to no effect of the chemical.

sum test suggests strong evidence against equality among the dose groups in maximum tumour burden per animal. Pairwise Wilcoxon tests were performed to test the equality of the maximum tumour burden between each treated group and the control, with one-sided alternatives of higher maximum burdens in the treated groups; the p-values turned out to be less than 0·01 for the three higher dose groups and were 1 and 0·09 for the 0·75 and 1·5 mg/kg groups, respectively. Similar results were obtained by taking the time of development of the first tumour as a summary of the tumour trajectory. As partly illustrated in Fig. 2, which shows the empirical and estimated mean tumour burdens for each group, the model provides a good fit to the data.

4·3. *Asthma inhaler use*

As a second application, we analysed data on daily usage of albuterol asthma inhalers (Grunwald et al., 2011). Daily counts of inhaler use were recorded over a period of 36 to 122 days for 48 students previously diagnosed with asthma. The total number of observations was 5209. As discussed in Grunwald et al. (2011), the data are underdispersed. Let y_{it} denote the number of times the ith student used the inhaler on day t. We are interested in the effect of morning levels of PM_{25}, small air pollutant particles of less than 25 mm in diameter, on asthma inhaler use. On each day t, a vector $x_t = (x_{t1}, \ldots, x_{tp})^{\mathrm{T}}$ of environmental variables was recorded, including PM_{25} level; average daily temperature, in degrees Fahrenheit, divided by 100; percentage humidity; and barometric pressure, in mmHg, divided by 1000. We modify (4) to include

these predictors in an additive model as follows:

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

where ξ_i is a random effect modelled as in § 4·2, $\{b_j\}$ is a B-spline basis, with the θ_j being the basis coefficients, and $\epsilon_i \sim N(0, \tau^{-1}R)$, with R being the correlation matrix arising from a first-order autoregressive process with correlation parameter ρ. The prior for each θ_j is identical to that used for θ_g in § 4·2, and each predictor is normalized to have zero mean and unit variance prior to analysis. The correlation parameter is given a uniform prior on $[-1, 1]$. The computational details are reported in the Supplementary Material.

We ran our Markov chain Monte Carlo algorithm for 11 000 iterations, discarding a burn-in of 1000 iterations. Convergence and mixing were diagnosed by monitoring the nonlinear effects of the different predictors at several values and also by monitoring hyperparameters. The trace plots showed excellent mixing, with an effective sample size of over 9000. Autocorrelation functions tended to drop near zero between lag 1 and lag 2. To obtain interpretable summaries of the nonlinear covariate effects on the inhaler use counts, we recorded for each predictor, at a dense grid of x_{jt} values for each sample after burn-in, the conditional expectation of the count for a typical student having $\xi_i = \mu_Q$:

$$\begin{aligned}\mu_j(x_{jt}) &= E(y_{it} \mid x_{jt}, x_{j't} = 0, j' \neq j,\ \xi_i = \mu_Q, \theta, \tau, \rho) \\ &\approx \sum_{k=0}^{\lceil K \rceil} k[\Phi\{a_{k+1}; \mu_j^*(x_{jt}), \tau\} - \Phi\{a_k; \mu_j^*(x_{jt}), \tau\}]. \end{aligned} \tag{5}$$

Here μ_Q is the mean of the random effects distribution Q, $\Phi(\cdot\,; \mu, \tau)$ 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

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

with the other predictors fixed at their mean value. Based on these samples, we calculated posterior means and pointwise 95% credible intervals; the results are reported in Fig. 4.

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

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

where u_i is a subject-specific random effect and e_{it} is a residual that follows a first-order autoregressive process. Grunwald et al. (2011) 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 confidence interval that is 50% wider. Our approach, which is based on a substantially more flexible model that allows for nonlinear effects and a nonparametric random effects distribution, produces results that are consistent with these earlier analyses.

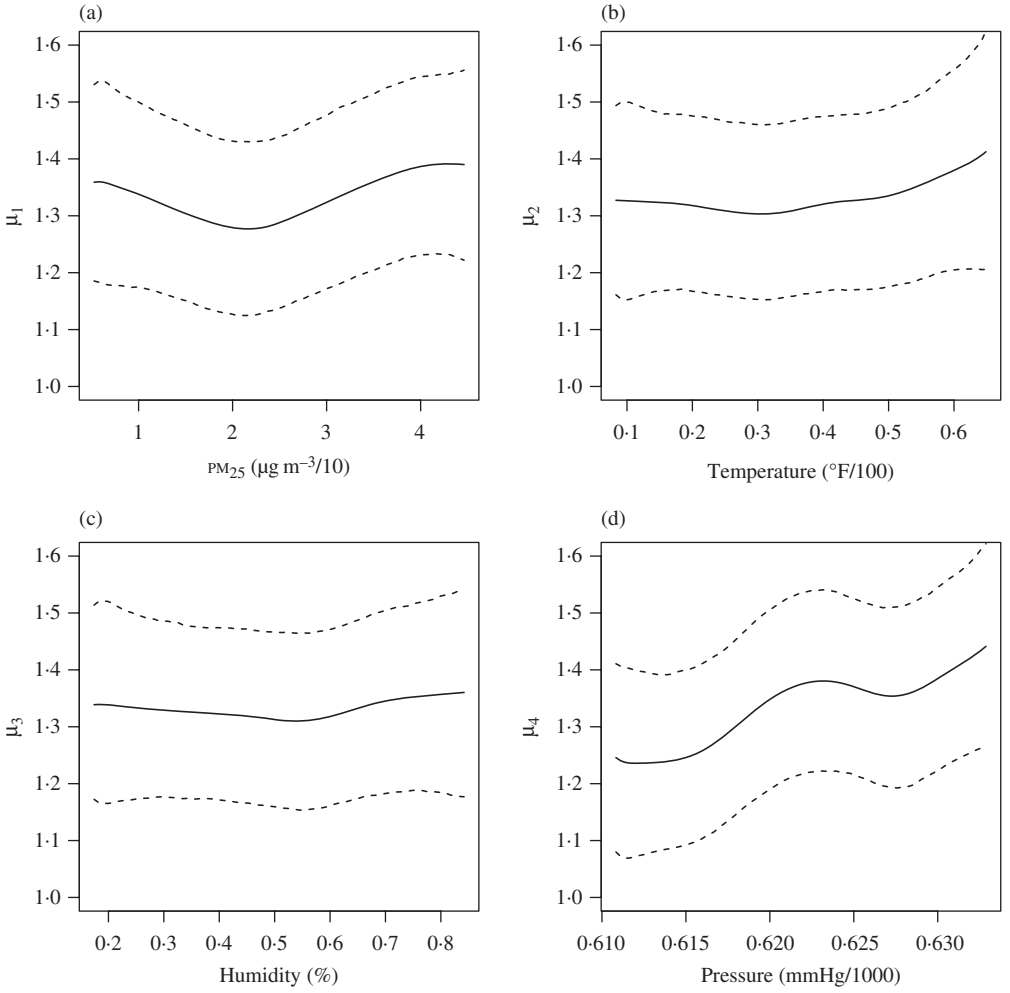

Fig. 4. Posterior mean (solid line) and 95% pointwise credible bands (dashed lines) for the effect of (a) concentration of PM_{25} pollutant, (b) average daily temperature, (c) percentage humidity, and (d) barometric pressure on asthma inhaler use calculated with equation (5).

5. Discussion

We have proposed a simple new approach to modelling 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 on count processes relies on Poisson processes 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 can be useful in this context. It is likely that some properties of the underlying continuous process, which are well known for Gaussian processes and other standard cases, will carry over to the induced count process. However, this issue merits further study. Other interesting directions to pursue include the modelling of counting processes corresponding to nondecreasing count processes via rounding nondecreasing continuous processes by

using monotone splines (Ramsay, 1998; Neelon & Dunson, 2004; Shively et al., 2009) and other constructions.

ACKNOWLEDGEMENT

Comments and suggestions of the referees are gratefully acknowledged. Antonio Canale is also affiliated with the Collegio Carlo Alberto. This research was partially supported by 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 of the covariance function of § 2·1, computational details for the rounded P-splines of § 2·3, an additional table and figure showing results of the simulation in § 3, and the computational details of the two applications discussed in § 4.

APPENDIX

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 in the interior of each $\mathcal{S}_l(y_0)$ and having unit increments at the boundaries. There are clearly infinitely many continuous functions $y^* : \mathcal{S} \to \mathbb{R}$ 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. The theorem is an immediate consequence of Lemma 1 and the following lemma which ensures that the mapping h maintains L^1 neighbourhoods.

LEMMA A1. *Suppose that y^* and y_0^* are continuous and bounded by $M \in \mathbb{R}$ and are such that $d_1(y^*, y_0^*) = \epsilon^*$. Let $y = h(y^*)$ and $y_0 = h(y_0^*)$. Then $y \in \eta_\epsilon(y_0)$ for all $\epsilon > \zeta(\epsilon^*; y_0^*)$, where $\zeta(\epsilon^*; y_0^*)$ is nondecreasing in ϵ^* and satisfies* $\lim_{\epsilon^* \to 0} \zeta(\epsilon^*; y_0^*) = 0$.

Proof of Lemma A1. Take $\mathcal{S} = [0, 1]^p$ without loss of generality. Let $\{\mathcal{S}_l(y_0, y)\}_{l=1}^m$ be 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 let $\lambda(\cdot)$ be Lebesgue measure. Define

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

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

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

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 a 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 $y^*(s) \leqslant a_{j_l+1}$ and $y_0^*(s) \leqslant a_{k_l+1}$. By construction $|a_{l_l+1} - a_{k_l+1}| \leqslant M_{\epsilon^*}$, and so for $M_{\epsilon^*} \to 0$ we have $a_{l_l+1} = a_{k_l+1}$. Considering that $\max |y^*(s) - y_0^*(s)| \to 0$, it follows that $|y^*(s) - y_0^*(s)| \to 0$ for all $s \in \mathcal{S}$, which leads also to $\max \delta_l \to 0$. As 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, then $\limsup_{\mathcal{S}} |y_0^*(s) - y^*(s)|$, and hence also M_{ϵ^*}, goes to zero as well.

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

By Lemma A1, 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 (2) as

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

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 become the whole space $\mathcal{C}$. It suffices to show that

$$\Phi_n \to 0, \tag{A1}$$

$$\exp(\beta_1 n) I_{1,n}(y_1, \ldots, y_n) \to 0, \tag{A2}$$

$$\exp(\beta_2 n) I_{2,n}(y_1, \ldots, y_n) \to 0, \tag{A3}$$

$$\exp(\beta n) I_{3,n}(y_1, \ldots, y_n) \to \infty, \tag{A4}$$

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

Denote by $\lfloor a \rfloor$ the integer part of a and let $\mathcal{S} = \bigcup_{j=1}^{\lfloor n^{1/p} \rfloor^p} \mathcal{G}_j$, with each $\mathcal{G}_j$ being an L^∞-ball of size $0{\cdot}5(\lfloor n^{1/p} \rfloor)^{-1}$ and centre s_j', where the centres 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 $(s_1, \ldots, s_n)^{\mathrm{T}}$ under Assumption 2. Define $X_i = 1\{y(s_i) = y_0(s_j')\}$, with s_j' being the centroid of the $\mathcal{G}_j$ which contains s_i. Let $\Phi_n = 1\{\sum_{i=1}^n X_i < n\}$ be the test on the set

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

where $M_n = O(n^\alpha)$ 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; hence, for n sufficiently large, the test functions have exactly zero Type I and Type II probability. From this (A1) is directly verified. We proceed to proving (A2). By Fubini's theorem we have

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

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

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

hence, for $\beta_2 < c_2$, $\exp(\beta_2 n) I_{2,n}(y_1, \ldots, y_n) \to 0$. Finally, the prior positivity of Π makes $I_{3,n}(y_1, \ldots, y_n)$ positive. This also establishes (A4) and thus completes 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. *J. R. Statist. Soc.* B **70**, 825–48.

BEHSETA, S., KASS, R. E. & WALLSTROM, G. L. (2005). Hierarchical models for assessing variability among functions. *Biometrika* **92**, 419–34.

CANALE, A. & DUNSON, D. B. (2011). Bayesian kernel mixtures for counts. *J. Am. Statist. Assoc.* **106**, 1528–39.

CHOI, T. & SCHERVISH, M. J. (2007). On posterior consistency in nonparametric regression problems. *J. Mult. Anal.* **98**, 1969–87.

CHU, W. & GHAHRAMANI, Z. (2005). Gaussian process for ordinal regression. *J. Mach. Learn. Res.* **6**, 1019–41.

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: Vine Copula Handbook*, Ed. D. Kurowicka & H. Joe, pp. 73–87. Singapore: World Scientific.

FADDY, M. J. (1997). Extended Poisson process modeling and analysis of count data. *Biomet. J.* **39**, 431–40.

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–30.

FRÜHWIRTH-SCHNATTER, S. & WAGNER, H. (2006). Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling. *Biometrika* **93**, 827–41.

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

GHOSAL, S. & ROY, A. (2006). Posterior consistency of Gaussian process prior for nonparametric binary regression. *Ann. Statist.* **34**, 2413–29.

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. *Biomet. J.* **53**, 578–94.

KACHOUR, M. & YAO, J. F. (2009). First-order rounded integer-valued autoregressive (RINAR(1)) process. *J. Time Ser. Anal.* **30**, 417–48.

LIECHTY, M. W. (2010). Multivariate normal slice sampling. *J. Comp. Graph. Statist.* **19**, 281–94.

MORRIS, J. S. & CARROLL, R. J. (2006). Wavelet-based functional mixed models. *J. R. Statist. Soc.* B **68**, 179–99.

MURRAY, I. & ADAMS, R. P. (2010). Slice sampling covariance hyperparameters of latent Gaussian models. In *Advances in Neural Information Processing Systems 23*, Ed. J. Lafferty, C. K. I. Williams, R. Zemel, J. Shawe-Taylor & A. Culotta, pp. 1723–31. New York: Curran Associates.

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. *J. Appl. Statist.* **37**, 1555–68.

RAMSAY, J. O. (1998). Estimating smooth monotone functions. *J. R. Statist. Soc.* B **60**, 365–75.

RODRÍGUEZ, A., DUNSON, D. B. & GELFAND, A. E. (2010). Latent stick-breaking processes. *J. Am. Statist. Assoc.* **105**, 647–59.

RUE, H., MARTINO, S. & CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Statist. Soc.* B **71**, 319–92.

SHIVELY, T. S., SAGER, T. W. & WALKER, S. G. (2009). A Bayesian approach to non-parametric monotone function estimation. *J. R. Statist. Soc.* B **71**, 159–75.

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. *J. R. Statist. Soc.* C **54**, 127–42.

VAN DER VAART, A. W. & VAN ZANTEN, J. H. (2009). Adaptive Bayesian estimation using a Gaussian random field with inverse gamma bandwidth. *Ann. Statist.* **37**, 2655–75.

WILSON, A. & GHAHRAMANI, Z. (2010). Copula processes. In *Advances in Neural Information Processing Systems 23*, Ed. J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. Zemel & A. Culotta, pp. 2460–68. New York: Curran Associates.

[*Received April* 2012. *Revised July* 2013]