Brazilian Journal of Probability and Statistics
2016, Vol. 30, No. 2, 155–170
DOI: 10.1214/14-BJPS268

A semiparametric Bayesian model for multiple monotonically increasing count sequences

Valeria Leiva-Yamaguchi and Fernando A. Quintana

Pontificia Universidad Católica de Chile

Abstract. In longitudinal clinical trials, subjects may be evaluated many times over the course of the study. This article is motivated by a medical study conducted in the U.S. Veterans Administration Cooperative Urological Research Group to assess the effectiveness of a treatment in preventing recurrence on subjects affected by bladder cancer. The data consist of the accumulated tumor counts over a sequence of regular checkups, with many missing observations. We propose a hierarchical nonparametric Bayesian model for sequences of monotonically increasing counts. Unlike some of the previous analyses for these data, we avoid interpolation by explicitly incorporating the missing observations under the assumption of these being missing completely at random. Our formulation involves a generalized linear mixed effects model, using a dependent Dirichlet process prior for the random effects, with an autoregressive component to include serial correlation along patients. This provides great flexibility in the desired inference, that is, assessing the treatment effect. We discuss posterior computations and the corresponding results obtained for the motivating dataset, including a comparison with parametric alternatives.

1 Introduction

Many longitudinal clinical trials arise when a time sequence of measurements is recorded for each of a number experimental units, allocated to one of several treatments. A key feature of longitudinal data that must be explicitly accounted for by realistic models, is the dependence of multiple responses obtained from the same individual. Frequently, the modeling effort is complicated by the presence of missing observations in the study. We specifically consider the case when the sequence of responses for each subject is integer-valued and monotonically increasing, that is, with nonnegative increments. The particular dataset that motivates our discussion comes from a U.S. Veterans Administration Cooperative Urological Research Group (VACURG) study about bladder cancer comparing the effectiveness of treatments to prevent the recurrence of Stage I bladder cancer (Byar, Blackard and Urological Research Group, 1977). Specifically, we focus on assessing the effect of thiotepa compared to a placebo. The data are available in Davis and Wei (1988), who considered a class of univariate one-sided global asymptotically distribution-free tests for the equality of the two treatments. Their testing and

Key words and phrases. Autoregressive model, Dirichlet process, generalized linear mixed model, hierarchical model.

Received July 2014; accepted November 2014.

estimation procedures assume that repeated measurements of the same characteristic, scheduled to be taken over a common set of time points for each study subject, are nondecreasing. Different forms of these data have been considered in the past. See, for example, Giardina et al. (2011), Di Lucca et al. (2013), and references therein. These works typically involve a dichotomization process that transforms each response in a (binary) tumor recurrence indicator, and with interpolation to deal with the missing responses. Unlike those approaches, we consider here the original accumulated tumor counts.

Our approach is built on generalized linear mixed models (GLMM) (e.g., Zeger and Karim, 1991; Breslow and Clayton, 1993) for longitudinal count variables, with an autoregressive structure to account for the serial correlation. We also explicitly consider the fact that the motivating dataset consists of cumulative counts of recurrent tumors, and therefore, the sequence of responses are nondecreasing over time. In the case of GLMMs, the distribution G of random effects is typically assumed to have a parametric distribution form, such as normal or some other heavier-tailed alternative. Such assumption may have critical impact on the type and quality of the inferences, specially if the random effects distribution is actually nonnormal and/or multimodal and/or skewed. We use instead a semiparametric approach (e.g., Ibrahim and Kleinman, 1998) where random effects are assumed a sample from a certain distribution G, in turn modeled by means of a Dirichlet process (DP) prior (Ferguson, 1973; Antoniak, 1974). Thus, individual heterogeneity is nonparametrically adjusted by random effects with a DP distribution. A remarkable stochastic representation of the DP as an infinite mixture of point masses was developed by Sethuraman (1994) (but see also Rolin, 1992):

$$G(B) = \sum_{h=1}^{\infty} w_h \delta_{\theta_h}(B), \qquad B \in \mathscr{B}, \tag{1}$$

where $w_1, w_2, \ldots$ are random weights given by $w_1 = V_1$ and $w_h = \prod_{j=1}^{h-1}(1 - V_j)V_h$ for $h > 1$, where $V_1, V_2, \ldots \overset{\text{i.i.d.}}{\sim} \text{Beta}(1, M)$, for some $M > 0$ called *total mass parameter*. In addition, $\theta_1, \theta_2, \ldots$ are a random sample (independent of the $\{w_h\}$ collection) from the *centering* distribution G_0, which is assumed to have support on a space that satisfies some minimal regularity conditions, and $\mathscr{B}$ is the Borel σ-field on the space where G_0 is supported. We denote this by $G \sim \mathscr{D}(M, G_0)$. In most applications, this space is Euclidean, which satisfies the technical requirements. An important consequence of (1) is that the DP is almost surely (a.s.) discrete. The support of the prior distribution is large, thus allowing for a wide range of shapes for G. This important property implies that by so modeling G, we obtain more flexible and reliable inferences than their parametric counterparts.

The DP has been very popular in Bayesian nonparametric statistics, playing a role similar to that of the normal distribution for parametric models. One key

reason for the popularity enjoyed by the DP is the availability of efficient posterior simulation schemes. At the heart of many of these algorithms is the result by Blackwell and MacQueen (1973) which states that for a random sample $X_1, \ldots, X_n | G \overset{\text{i.i.d.}}{\sim} G$ and $G \sim \mathscr{D}(M, G_0)$, the joint marginal distribution of $X_1, \ldots, X_n$ can be expressed as a sequence of draws from a Pólya urn with a continuum of colors, distributed as G_0. For the purpose of posterior simulation, the most relevant consequence is the implied predictive distribution:

$$p(X_i | \mathbf{X}_{-i}) \propto \sum_{j=1,\, j\neq i}^{n} \delta_{X_j}(X_i) + M G_0(X_i), \qquad i = 1, \ldots, n, \tag{2}$$

where $\mathbf{X}_{-i}$ represents the vector $(X_1, \ldots, X_n)$ with the ith coordinate removed, and the proportionality constant is $M + n - 1$. A colorful description of this sequence of predictive distributions has been termed the Chinese restaurant process (Aldous, 1985). The distribution in (2) gives also an explicit interpretation of the discreteness property of DPs pointed out earlier. See further details on posterior simulation for DP-based models in for example, Escobar (1994), Bush and MacEachern (1996), MacEachern and Müller (1998), and Neal (2000), among many other references. A different type of algorithm that works directly with (1) can be found in Ishwaran and James (2001). Review of models based on the DP and many other alternatives can be found in Müller and Quintana (2004), Hjort et al. (2010) and in Müller and Mitra (2013).

The rest of the paper is organized as follows. Section 2 describes the data and motivates the model discussed in Section 3, where we also emphasize its particular likelihood and prior components. Section 4 presents the main results of the inferences on quantities of interest. In particular, we find a significant treatment effect, in the sense that treated patients have smaller bladder tumor recurrence. As part of the analysis, we also assess sensitivity of the results to various prior choices. A comparison with a parametric alternative is considered as well. Furthermore, we focus on predictions for future observations of current patients or new patients either under treatment with thiotepa or under placebo. Section 5 presents a final summary and conclusions from the analysis. Some computational details are presented in an Appendix.

2 Data description

Due to the high recurrences rate and the need for lifelong care and monitoring, bladder cancer is the most expensive cancer to treat on a per-patient basis. If a bladder cancer only affects the inner lining of the bladder, it is known as a superficial cancer or Stage I bladder cancer. Stage I tumors can usually be completely removed by transurethral resection, but many patients have multiple recurrence. The subsequent tumors sometimes show a higher degree of malignancy and may even progress to invasive carcinoma.

The data we consider came from a randomized clinical trial conducted by the VACURG, concerning the effect of drug thiotepa on the recurrence of tumors in the bladder. Thiotepa is a chemotherapy drug used to reduce the size of a cancerous tumors and prevent the growth of a new cancer cell. At the beginning of the trial, all available patients had superficial bladder tumors. To determine if recurrence of Stage I bladder cancer can be prevented, the tumors were removed through the urethra and patients were assigned to one of two treatments: placebo and thiotepa. At subsequent follow up visits, all recurrent tumors were removed and treatment continued.

The data, sequences of accumulated tumor counts at subsequent follow-up checks every three months, are presented in Davis and Wei (1988). A total of 82 patients were assigned to thiotepa (46) and placebo (36) treatments. The number of tumors was measured at the initial time and at each of the follow-up visits every three months (3 to 36 months). Each sequence is thus monotonically increasing, despite the fact that there are many missing observations, because not all patients showed up at every scheduled time.

Our main aim is to determine whether usage of thiotepa decreases the recurrence of new tumors over time. To do so, and considering the nondecreasing nature of the observations, we work with the count difference over two consecutive visits. Thus, let y_{ij} denote the increase in the number of tumors between time $j-1$ and j for patient i (we define $y_{i0}=0$), with $j=1,\ldots,n_i$, where n_i is the corresponding number of measurements. As in Davis and Wei (1988), we will assume that the missing value process is independent of the observations. We also let $x_i=1$ if patient i was assigned to thiotepa treatment, and $x_i=0$ otherwise (i.e., placebo group).

3 Model formulation

To deal with the data features and to assess treatment effect, we here adopt a GLMM formulation. Specifically, we assume a Poisson distribution for each tumor count difference y_{ij} with rate λ_{ij}. To include correlation along the sequences, we assume an autoregressive structure, where the distribution of y_{ij} depends on the previous $y_{i,j-1}$. We do so by including a linear term in the link function, that is, $\log(\lambda_{ij})$ includes a linear transformation of $y_{i,j-1}$. In addition, individual heterogeneity is captured by an additive random effect b_i, which is modeled nonparametrically with a Dirichlet process prior.

Our hierarchical model formulation is then given by

$$\begin{aligned} y_{ij}|\lambda_{ij}, y_{i,1:j-1} &\sim \text{Poisson}(\lambda_{ij}),\\ \log(\lambda_{ij}) &= \alpha_0+\alpha_1 x_i+(\beta_0+\beta_1 x_i)y_{i,j-1}+b_i,\\ b_i|G &\sim G, \qquad G\sim \mathscr{D}(M,G_0), \qquad G_0|\tau\sim N(0,\tau),\\ \alpha_0,\alpha_1,\beta_0,\beta_1 &\sim N(0,B), \qquad \tau^{-1}\sim \text{Gamma}(\tau_a,\tau_b), \end{aligned} \tag{3}$$

where $y_{i,1:j-1} = (y_{i,1}, \ldots, y_{i,j-1})$, and B, τ_a and τ_b are known positive constants specified by the user. Here, B is the prior variance of the coefficients, and the Gamma distribution is parametrized so that the precision τ^{-1} has prior mean $E(\tau^{-1}) = \tau_a/\tau_b$. Note that the likelihood is defined as a sequence of increasing conditionals that convey dependence only on the previous element in the sequence. Since there are four regression parameters in the likelihood model, we follow the convention in Giardina et al. (2011) and refer to model (3) as BNP4P.

The likelihood in BNP4P model (3) corresponds to a Poisson regression, where the linear predictor involves an interaction between the treatment indicator variable x_i and the first lagged increment $y_{i,j-1}$. This interaction is formalized by the β_1 term in the likelihood model, which allows for differences in how the log of the mean tumor count increments change as time progresses for treated compared to placebo patients. This is similar in spirit to the models in Giardina et al. (2011) and Di Lucca et al. (2013), albeit with different outcome variables and likelihood model. Generally speaking, we can interpret the linear predictor coefficients in terms of the log-expected counts. Thus, parameters (α_0, α_1) describe the average effect of the corresponding treatment received, while (β_0, β_1) describe the effect of each patient deviation from the mean of treatment subject to the number of tumors. The DP assumption on the b_i random effects allows for shapes beyond the usual normal assumption. Because of the Sethuraman (1994) representation discussed earlier, there may be ties in the b_i values, leading to the well-known clustering property of the DP (and, in fact, of any discrete random probability measure). As a consequence, patients can be clustered by the value of their random effects, while still retaining their particular longitudinal evolution. The centering distribution is a zero-mean normal distribution with variance τ, given itself a standard inverse-Gamma distribution, which allows for adaptation of the centering distribution, as required by the data.

As a natural simplified version of model BNP4P, we consider setting β_1 to zero in the likelihood in (3). In other words, under the simplified version, the log-transformed mean function then becomes

$$\log(\lambda_{ij}) = \alpha_0 + \alpha_1 x_i + \beta_0 y_{i,j-1} + b_i. \tag{4}$$

Following again the notation in Giardina et al. (2011), we refer to (4) as the three-parameter model (BNP3P). Under model BNP3P the interaction between treatment indicator and lagged responses commented above is suppressed, and this lagged term affects only the baseline mean increment in the number of tumors.

It is also important to point out that we assume the mechanism provoking the missing values to be independent of the actual data and that these unobserved values follow the same model. We will assume this is the case for all the models considered in our analysis. To deal with missing values we therefore use an imputation scheme in the MCMC algorithm discussed later in Section 4. As a result, the inference will be based on the posterior distribution after marginalization over the missing values by averaging over the corresponding imputations.

4 Data analysis

4.1 Computation

Model (3) implies an analytically intractable posterior distribution and therefore, the need to use MCMC methods to simulate from it. As usual in DP-based models, we introduce cluster indicators to facilitate the simulation. Recall that the discreteness of the DP implies that there are ties among $b_1, \ldots, b_n$. Let $b_1^*, \ldots, b_k^*$ denote the unique values among $b_1, \ldots, b_n$. A partition of the entire collection of patients is thus induced by identifying these unique values. Let $S_1, \ldots, S_k$ denote the nonempty subsets in the partition, and let $c_i = j$ if $i \in S_j$, $j = 1, \ldots, k$, so that $b_i = b_{c_i}^*$ for $i = 1, \ldots, n$. The entire $b_1, \ldots, b_n$ collection is then equivalent to $b_1^*, \ldots, b_k^*$ and $c_1, \ldots, c_n$.

The most time-consuming step of the MCMC algorithm corresponds to the updating of the configurations $c_1, \ldots, c_n$. In our case, we implemented Algorithm 8 in Neal (2000), which is particularly suitable to the fact that the likelihood and the centering distribution are nonconjugate. In addition, the corresponding full conditionals for the autoregressive coefficients are not available in closed form, and therefore, we used adaptive methods as discussed in Roberts and Rosenthal (2009). See details in the Appendix.

We finally point out that only for the purpose of initializing our Markov chain, we imputed each missing value in the cumulative counts by a linear interpolation.

4.2 Results

We ran the posterior simulation scheme using 100,000 scans of which 12,000 were burned, with a thinning of 50. We judged practical convergence as determined by standard tests such as those contained in the CODA package (Plummer et al., 2006). We chose $B = 10^3$ and $\tau_a = \tau_b = 10^{-3}$, implying vague prior specifications, thus reflecting a genuine lack of prior information. Summaries of the marginal posterior distributions for α_0, α_1, β_0, β_1 and τ are shown in Table 1. The corresponding posterior densities are presented in Figure 1.

Interestingly, the posterior distributions of α_0, α_1 and β_1 are mostly concentrated on the negative numbers, and on the positive numbers for β_0. Looking at the corresponding posterior supports, it follows that on average, the initial number of tumors is less than one, and even less for the treated group. In addition, the increase in number of tumors over consecutive follow-up visits is, on the average, less than one as well. There is also some evidence that this increase is even less for the thiotepa group, because $P(\beta_1 < 0|\text{data}) = 0.58523$.

Figure 2 shows predictions for an entire sequence of twelve measurements for one new patient from each of the placebo and treatment groups. In other words, we predict outcomes for future patients. The displayed curves correspond to the respective posterior predictive means. As a comparison, we also include the same

Table 1 *Marginal posterior summaries for all the models considered in this work, including posterior mean, median, standard deviation and* 95% *highest posterior density (HPD) intervals. The left panel shows results for the nonparametric models, while the right one is for the corresponding parametric versions. The upper segment shows results for the BNP4P and BP4P models; the middle segment is dedicated to BNP3P and BP3P; and the lower segment is for the BPBM model given in* (5)*, that is, the version of BP4P with group-specific variances*

	Nonparametric				Parametric			
	Mean	Median	Std. Dev.	HPD	Mean	Median	Std. Dev.	HPD
α_0	−1.73	−1.71	0.30	(−2.34; 1.16)	−1.63	−1.63	0.32	(−2.29; −0.097)
α_1	−0.98	−0.98	0.09	(−1.17; −0.79)	−1.13	−1.14	0.48	(−2.13; −0.133)
β_0	0.10	0.10	0.02	(0.05; 0.15)	−0.038	−0.037	0.326	(−0.102; 0.023)
β_1	−0.05	−0.05	0.05	(−0.15; 0.06)	−0.21	−0.21	0.058	(−0.32; −0.099)
τ	3.06	2.36	2.25	(0.46; 7.43)	0.31	0.30	0.09	(0.151; 0.499)
α_0	−1.90	−1.87	0.28	(−2.46; −1.37)	−1.65	−1.646	0.356	(−2.31; 0.99)
α_1	−0.35	−0.36	0.19	(−0.75; 0.03)	−1.05	−1.041	0.51	(−2.091; −0.08)
β_0	0.16	0.16	0.02	(0.11; 0.21)	0.018	0.019	0.026	(−0.033; 0.07)
τ	2.3	1.75	2.22	(0.33; 5.59)	0.306	0.2955	0.088	(0.151; 0.184)
α_0					−1.611	−1.587	0.357	(−2.33; −0.925)
α_1					−1.158	−1.12	0.641	(−2.44; 0.08)
β_0					−0.038	−0.037	0.032	(−0.103; 0.023)
β_1					−0.213	−0.213	0.058	(−0.32; −0.09)
τ_1					0.314	0.297	0.112	(0.123; 0.538)
τ_2					0.351	0.3164	0.178	(0.072; 0.705)

estimated curves, obtained using a parametric model that follows when letting $M \to \infty$, that is, assuming $b_1, \ldots, b_n$ to be distributed according to the baseline distribution G_0. By analogy to the proposed model, we refer to this parametric alternative as BP4P, and to the parametric counterpart of (4) as BP3P. All the curves are monotonic, as to be expected by construction of the model. It also follows, for both models, that the predicted mean curves for the placebo patient are uniformly above the curve for the treated patient, revealing that there is indeed a treatment effect. Furthermore, the placebo curve grows at a higher rate than the treated curve, which simply reflects the fact that β_1 has slightly more of its posterior mass on the negative numbers (see above). Interestingly, the parametric model predicts curves that have essentially identical start but that deviate from each other after the third occasion. In contrast, the proposed model predicts curves that are separated from the start, thus uncovering an early treatment effect that increases over time.

Figure 3 shows a smoothed version of the posterior predictive distribution of a new random effect under each model. By definition, the parametric model is too constrained by the normality assumption, while the proposed nonparametric model clearly exhibits a bimodal feature. We further investigate this issue in Section 4.3.

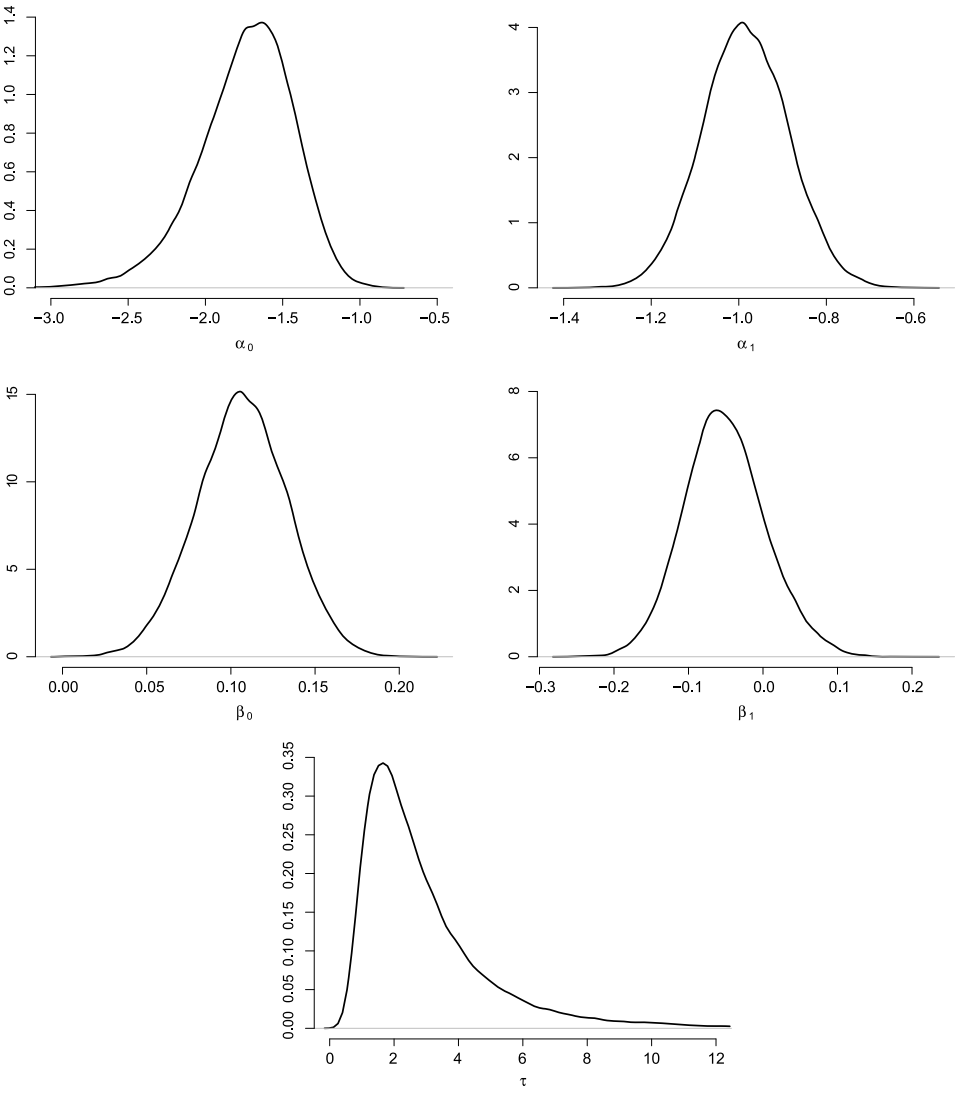

Figure 1 *Marginal posterior distribution for* α_0, α_1, β_0, β_1, *and* τ *under model BNP4P.*

4.3 Sensitivity analysis and other comparisons

In addition to the analysis presented so far, we also explored sensitivity of the results to various prior specifications. We considered varying M to be one of $\{2, 3, 4, 5\}$, keeping all the other prior definitions as earlier. The results were essentially the same as those reported so far. Next, we tried changing the priors for τ and the regression coefficients to $B = 10$ and $\tau_a = 10^{-2}$ and $\tau_b = 10^{-5}$. Again, the results changed very little, and in particular, the predictions were undistinguishable from those in Figure 2, and so we do not report them.

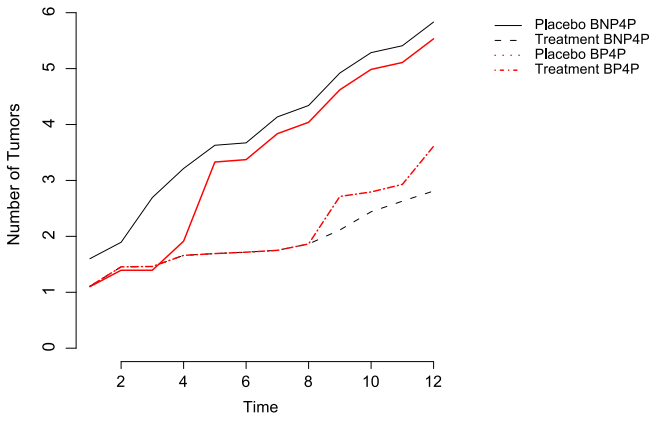

Figure 2 *Estimated posterior predictive mean at each follow-up occasion for one thiotepa group new patient and one placebo group new patient under the proposed nonparametric model (BNP4P) and under a parametric version of the model (BP4P). The curves are connected for the purpose of presentation only.*

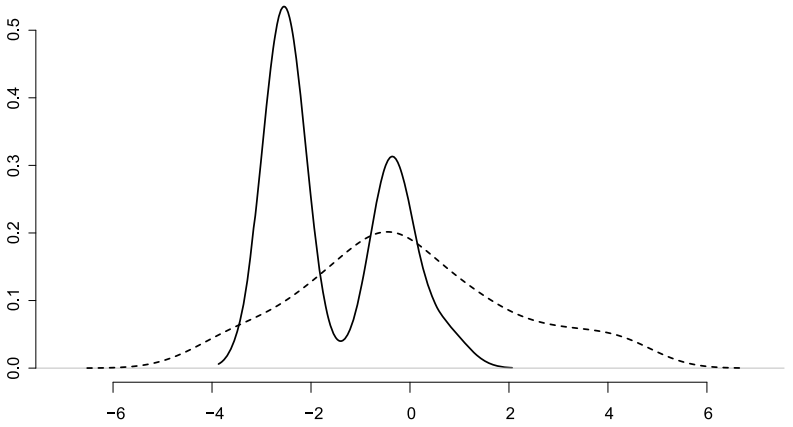

Figure 3 *Estimated posterior density of b under the parametric alternative BP4P (dotted line), together with posterior predictive density of a new random effect b_{n+1} under the proposed nonparametric model BNP4P (solid line).*

Turning now our attention to the interaction between treatment and lagged responses, we compare models BNP4P and BNP3P. The MCMC details for BNP3P are almost identical to those of BNP4P and therefore we omit them. Table 1 shows posterior summaries obtained under model BNP3P. The numerical values change somehow, but the general conclusion remains the same, namely, that there is indeed an average reduction in the incremental number of tumors over time when using the thiotepa drug. In fact, this is clearly shown in the predicted curves for a new treated and a new placebo patient, exhibited in Figure 4. The predictions clearly

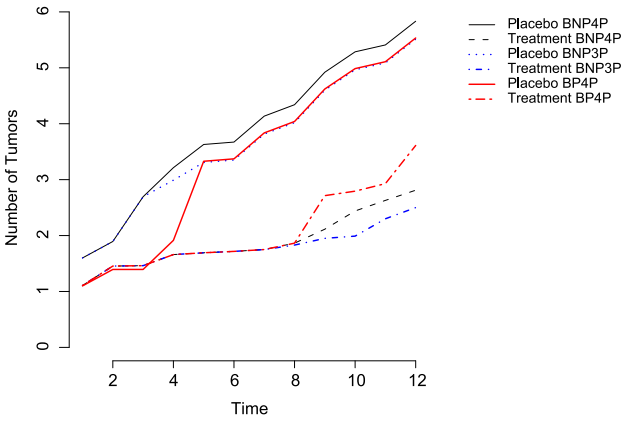

Figure 4 *Comparison of predictive distribution under models BNP3P, BNP4P and the* 4*-parameter parametric version* (*BP4P*).

depict a difference between patients in placebo and treatment groups. Starting from about the same values, it is soon evident that thiotepa treatment gets a considerable smaller expected variation in the accumulated number of tumors, encoding what we have learned from data. Interestingly, the same general conclusion can be achieved using the parametric alternatives BP4P and BP3P. Posterior summaries for each of these are shown in Table 1, and predictions for the former are included in Figure 4. Although these curves are not identical to each other within each of the two groups, they show similar trends. In all cases, the predicted ratio of expected number of tumors for placebo to treatment grows from about 1.5 at baseline to about 2.0 after the twelfth visit (36 months). This represents a substantial gain from the use of thiotepa.

A further comparison of the parametric and nonparametric models discussed so far can be stated in terms of the log pseudo marginal likelihood (LPML) and the deviance information criterion (DIC). The LPML was introduced by Geisser and Eddy (1979) as a measure of predictive accuracy. Gelfand and Dey (1994) suggested a particularly simple way to carry out the required computations from MCMC output. Having several competing models, the one achieving the highest LPML is to be preferred. On the other hand, the DIC was introduced by Spiegelhalter et al. (2002) and later discussed by Celeux et al. (2006) for finite mixture and random effects models. Models with lower DIC values are to be preferred.

Table 2 shows the LPML and the DIC criteria for each of the nonparametric models BNP3P and BNP4P and the parametric models BP4P and BP3P. Both criteria find that the nonparametric models are to be preferred to the parametric ones. They also show that BNP3P has a somewhat better performance than BNP4P. In other words, the increased complexity of adding the interaction parameter in BNP4P compared to BNP3P does not imply a substantial improvement in model

Table 2 *Comparison of parametric and semiparametric models*

Model	LPML	DIC
BN4P	−639.52	1198.37
BNP3P	−629.69	1191.13
BP4P	−725.42	1285.26
BP3P	−746.63	1342.64
BPBM	−733.02	1315.72

fitting, according to these criteria. This is also reflected in the rather small differences in the predictions shown in Figure 4. However, a small interaction effect can still be found in these predictions, as the BNP4P and BNP3P curves do not have exactly the same growth rates. Indeed, the corresponding placebo and treatment curves appear to deviate from each other at different places. We interpret this situation by concluding that an interaction is present but it is not strong enough to imply an improvement over the penalty related to increased model complexity. We point out that BNP3P has no built-in interaction term, and so the mean structure would not reflect this, even if such effect was really present.

Finally, looking back at the bimodal shape of the predictive distribution depicted in Figure 3, one may naively think that it is due to the presence of different variance components for treated and placebo patients. To further investigate this issue, we consider now yet another parametric variation of model BNP4P, where $b_1, \ldots, b_n$ are independent, and

$$b_i \sim \begin{cases} N(0, \tau_1), & \text{if } x_i = 1, \\ N(0, \tau_2), & \text{if } x_i = 0. \end{cases} \tag{5}$$

We refer to model (5) as BPBM because of the bimodality that motivated it. The corresponding LPML and DIC values are indicated in Table 2. Again, the two semiparametric alternatives prevail. Interestingly, BPBM is outperformed by BP4P for which there is a single variance component τ rather than two. So, the bimodality is really not explained by two specific variance components related to the treatments, but actually due the fact that the nonparametric distribution of BNP3P and BNP4P, by the Sethuraman representation (1), supports an infinite number of components, and ties among these components have a distribution described by the Pólya urn representation with predictive distribution (2).

5 Summary

We proposed a model for multiple longitudinal sequences of nondecreasing counts. Our approach was based on generalized linear mixed models, with a nonparametric model for random effects. The model was motivated by data coming from a

clinical study of the effect of thiotepa on Stage I bladder cancer, with multiple missing values, assumed to arise from a process independent from the observations. To account for longitudinal correlation and monotonicity, we proposed an autoregressive Poisson model on the subsequent tumor increments. The link function includes a 4-parameter formulation that allows for treatment interactions. Our results suggest that there is indeed a positive effect of thiotepa in the sense that, on the average, treated patients had tumor recurrence at a slower rate than those assigned to the placebo group.

We studied sensitivity of the results to various prior specifications, but the overall conclusions remained the same under all the cases we considered. We also compared our results with a parametric formulation of the model, as well as with a simplified 3-parameter version, where one of the parameters from the original model was eliminated. The proposed model clearly outperforms the parametric alternative, but the 3-parameter alternative has a slightly better fit to the data, as determined by applying model comparison criteria such as LPML and DIC. Nevertheless, the proposed model has the advantage of including an explicit interaction between treatment and the autoregressive component, thus allowing for individual tumor growth rates that differ among groups. This is a conceptually important difference, specially for predictive inference.

Appendix: Computational details

We now give details on Markov chain Monte Carlo simulation to generate posterior draws for the proposed model. Details are specific to the BNP4P model, but a slight modification can be used to implement similar calculations for the BNP3P model. The steps to be described next were implemented using C.

Updating the configurations and cluster-specific parameters

The most time-consuming step in our Gibbs sampling implementation of posterior simulation for the proposed BNP4P model is the updating of the clustering structure. Recall that the cluster membership indicators $c_1, \ldots, c_n$ are such that $b_i = b^*_{c_i}$ for $i = 1, \ldots, n$, where $b^*_1, \ldots, b^*_k$ are the k unique values among $b_1, \ldots, b_n$, that is, k is the number of clusters. To update the configurations, we use Algorithm 8 in Neal (2000) with $m = 1$. This algorithm is particularly useful when the likelihood and the baseline distribution are not in conjugate form. This is our case, as the likelihood contribution of the ith patient is

$$L_i(\boldsymbol{\eta}, b_i|\mathbf{y}_i) = \prod_{j=1}^{n_i} \text{Po}(y_{ij}|\lambda_{ij}), \tag{A.1}$$

where $\text{Po}(y|\lambda) = \lambda^y e^{-\lambda}/y!$ is the Poisson probability mass function for $\lambda > 0$, $\boldsymbol{\eta} = (\alpha_0, \alpha_1, \beta_0, \beta_1)$, λ_{ij} was defined in (3), and the baseline distribution is Gaussian.

To implement the algorithm by Neal (2000), we introduce m temporary additional parameters $b^*_{k+1}, \ldots, b^*_{k+m}$ that are associated to "empty clusters" that are to be discarded when not used. Let k^- the number of distinct c_j's for $j \neq i$, and let n^-_ℓ the number of c_j's equal to ℓ for $j \neq i$. Let $h = k^- + m$ and label the c_j's for $j \neq i$ in $\{1, \ldots, k^-\}$. If observation i was in a singleton, assign c_i the label $k^- + 1$ and draw independent values $b^*_{k^-+2}, \ldots, b^*_h$ from the baseline distribution G_0. Then resample c_i from

$$P(c_i = \ell | \text{everything else}) \propto \begin{cases} n^-_\ell L_i(\boldsymbol{\eta}, b^*_\ell | \mathbf{y}_i), & \text{for } 1 \leq \ell \leq k^-, \\ \dfrac{M}{m} L_i(\boldsymbol{\eta}, b^*_\ell | \mathbf{y}_i), & \text{for } k^- + 1 \leq \ell \leq h, \end{cases}$$

where the likelihood L_i is given in (A.1). This step is repeated for $i = 1, \ldots, n$. See further details in Neal (2000). For our particular implementation, we use $m = 1$.

With the new configurations available, the number of clusters is known, say k, and the unique values are updated from

$$p(b^*_\ell | \text{else}) \propto \prod_{i\,:\,c_i=\ell} \prod_{j=1}^{n_i} \text{Po}(y_{ij} | \lambda^*_{ij}(\ell)) \times \exp\left(-\frac{b^{*2}_\ell}{2\tau}\right),$$

where $\lambda^*_{ij}(\ell) = \exp(\alpha_0 + \alpha_1 x_i + (\beta_0 + \beta_1 x_i) y_{i,j-1} + b^*_\ell)$. This is accomplished by resorting to a Metropolis-within Gibbs step with a random walk-type of chain.

Updating the autoregressive and baseline parameters

Each of the four autoregressive and the baseline distribution parameters are updated using the adaptive Metropolis–Hastings algorithm described in Roberts and Rosenthal (2009). In words, the algorithm considers a random walk-type of Metropolis–Hastings procedure with a zero-mean candidate generator, and a variance σ^2 that is changed along the way. The basic rationale is to target a rejection rate of 44%. To do so, we consider batches of a certain number of chain elements (50, say), and either increase or decrease σ^2 by δ_r depending on whether the rejection rate in the rth batch was more or less than 44%. Following the suggestion by Roberts and Rosenthal (2009), we choose $\delta(r) = \min\{0.01, r^{-1/2}\}$.

The above idea is applied separately to each of the four autoregressive parameters α_0, α_1, β_0 and β_1. Let

$$\lambda^*_{ij}(\ell) = \exp(\alpha_0 + \alpha_1 x_i + (\beta_0 + \beta_1 x_i) y_{i,j-1} + b^*_\ell). \tag{A.2}$$

Then, the corresponding conditional distributions to which this method is applied to, are given by

Conditional distribution of α_0:

$$p(\alpha_0 | \text{else}) \propto \prod_{\ell=1}^{k} \prod_{i\,:\,c_i=\ell} \prod_{j=1}^{n_i} \lambda^*_{ij}(\ell)^{y_{ij}} \exp(-\lambda^*_{ij}(\ell)) \exp\left(-\frac{\alpha_0^2}{2B}\right).$$

Conditional distribution of α_1:

$$p(\alpha_1|\text{else}) \propto \prod_{\ell=1}^{k} \prod_{i:\, c_i=\ell} \prod_{j=1}^{n_i} \lambda_{ij}^*(\ell)^{y_{ij}} \exp\bigl(-\lambda_{ij}^*(\ell)\bigr) \exp\biggl(-\frac{\alpha_1^2}{2B}\biggr).$$

Conditional distribution of β_0:

$$p(\beta_0|\text{else}) \propto \prod_{\ell=1}^{k} \prod_{i:\, c_i=\ell} \prod_{j=1}^{n_i} \lambda_{ij}^*(\ell)^{y_{ij}} \exp\bigl(-\lambda_{ij}^*(\ell)\bigr) \exp\biggl(-\frac{\beta_0^2}{2B}\biggr).$$

Conditional distribution of β_1:

$$p(\beta_1|\text{else}) \propto \prod_{\ell=1}^{k} \prod_{i:\, c_i=\ell} \prod_{j=1}^{n_i} \lambda_{ij}^*(\ell)^{y_{ij}} \exp\bigl(-\lambda_{ij}^*(\ell)\bigr) \exp\biggl(-\frac{\beta_1^2}{2B}\biggr).$$

Conditional distribution of τ: For τ we have the following closed form full conditional:

$$\tau^{-1}|\text{else} \sim \text{Gamma}\biggl(\tau_a + \frac{k}{2}, \tau_b + \frac{1}{2}\sum_{\ell=1}^{k} b_\ell^{*2}\biggr).$$

Missing values

Finally, every iteration of the MCMC requires imputing the missing observations, which we do sequentially. Given the currently imputed parameter values, then for $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$, if $y_{i,j}$ was missing, we impute a value by drawing $y_{ij} \sim \text{Po}(\lambda_{ij}^*(\ell))$, where the $\lambda_{ij}^*(\ell)$ quantity was defined in (A.2). Note that (A.2) can be readily evaluated as $y_{i,j-1}$ is either an actual observation or a previously imputed missing observation.

Acknowledgment

This research was partially funded by Grant FONDECYT 1100010.

References

Aldous, D. J. (1985). *Exchangeability and Related Topics, École d'été de probabilités de Saint-Flour, XIII—1983*. *Lecture Notes in Math.* **1117**, 1–198. Berlin: Springer. MR0883646

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Annals of Statistics* **2**, 1152–1174. MR0365969

Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via Pólya urn schemes. *Annals of Statistics* **1**, 353–355. MR0362614

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized in linear mixed models. *Journal of the American Statistical Association* **88**, 9–25.

Bush, C. A. and MacEachern, S. N. (1996). A semiparametric Bayesian model for randomised block designs. *Biometrika* **83**, 275–285.

Byar, D., Blackard, C. and Urological Research Group (1977). Comparisons of placebo, pyridoxine, and topical thiotepa in preventing recurrence of stage I bladder cancer. *Urology* **10**, 556–561.

Celeux, G., Forbes, F., Robert, C. P. and Titterington, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis* **1**, 651–673 (electronic). MR2282197

Davis, C. S. and Wei, L. J. (1988). Nonparametric methods for analyzing incomplete nondecreasing repeated measurements. *Biometrics* **44**, 1005–1018. MR0980996

Di Lucca, M., Gugliemi, A., Müller, P. and Quintana, F. A. (2013). A simple class of Bayesian nonparametric autoregression models. *Bayesian Analysis* **8**, 63–88. MR3036254

Escobar, M. D. (1994). Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association* **89**, 268–277. MR1266299

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics* **1**, 209–230. MR0350949

Geisser, S. and Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association* **74**, 153–160. MR0529531

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B, Methodological* **56**, 501–514. MR1278223

Giardina, F., Guglielmi, A., Ruggeri, F. and Quintana, F. A. (2011). Bayesian first order autoregressive latent variable models for multiple binary sequences. *Statistical Modelling International Journal* **11**, 471–488.

Hjort, N. L., Holmes, C., Müller, P. and Walker, S. (2010). *Bayesian Nonparametrics*. Cambridge, UK: Cambridge University Press. MR2722987

Ibrahim, J. G. and Kleinman, K. P. (1998). Semiparametric Bayesian methods for random effects models. In *Practical Nonparametric and Semiparametric Bayesian Statistics*. *Lecture Notes in Statist.* **133**, 89–114. New York: Springer. MR1630077

Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* **96**, 161–173. MR1952729

MacEachern, S. N. and Müller, P. (1998). Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics* **7**, 223–338.

Müller, P. and Mitra, R. (2013). Bayesian nonparametric inference—Why and how. *Bayesian Analysis* **8**, 269–302. MR3066939

Müller, P. and Quintana, F. A. (2004). Nonparametric Bayesian data analysis. *Statistical Science* **19**, 95–110. MR2082149

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* **9**, 249–265. MR1823804

Plummer, M., Best, N., Cowles, K. and Vines, K. (2006). Coda: Convergence diagnosis and output analysis for MCMC. *R News* **6**, 7–11.

Roberts, G. O. and Rosenthal, J. S. (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics* **18**, 349–367. MR2749836

Rolin, J.-M. (1992). Some useful properties of the Dirichlet process. Technical Report 9207, Center for Operations Research & Econometrics, Univ. Catholique de Louvain.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 639–650. MR1309433

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B, Statistical Methodology* **64**, 583–639. MR1979380

Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effects; a Gibbs sampling approach. *Journal of the American Statistical Association* **86**, 79–86. MR1137101

Departamento de Estadística
Pontificia Universidad Católica de Chile
Avenida Vicuña Mackenna 4860
Macul, Santiago
Chile
E-mail: vjleiva@mat.puc.cl
quintana@mat.uc.cl