Modelling correlated binary variables: an application to lower urinary tract symptoms

William Barcella* and Maria De Iorio

Department of Statistical Science, University College London, UK

James Malone-Lee

Department of Medicine, University College London, UK

February 24, 2018

Abstract

We present a semi-parametric model for time-evolving vectors of correlated binary variables. We introduce continuous latent variables which are discretised to obtain the sampling model. We assume the distribution of the latent variables to be an infinite mixture of distributions with weights that vary across some covariate space and with mean and covariance matrix being component-specific. This distribution includes also an autoregressive term that captures the time evolution of the latent variables and therefore of the binary observations. The proposed method is motivated by the study of lower urinary tract symptoms observed at subsequent attendance visits. In particular, we evaluate the temporal dependence among the symptoms controlling for the presence of urinary tract infection. The results show that the most recurrent symptoms are stress incontinence and voiding, which are also the most related with presence of pyuria, the best biomarker of infections. Furthermore, we observe that the correlation among symptoms changes over time. The pair of symptoms which appear to be the most correlated are pain and voiding.

Keywords: Bayesian nonparametric regression; Dependent generalised Dirichlet process; Correlated binary variables; Dynamic Probit model; Multivariate Probit model.

*william.barcella.13@ucl.ac.uk

1 Introduction

Lower urinary tract symptoms (LUTS) represent a group of signs which appear as the consequence of a number of possible diseases: from neurological problems, to anxiety and *urinary tract infections* (UTI). These symptoms are commonly classified into four categories: urgency symptoms, sudden urge to urinate; stress incontinence symptoms, episodes of incontinence caused by stressing the bladder; voiding symptoms, problems in voiding the bladder; and pain symptoms, pain while urinating or when the bladder is full. LUTS affect a large proportion of the population, especially elderly people, and they contribute significantly to the costs of the health systems. In fact, the diseases that lead to the presence of LUTS can often become chronic, as such requiring expensive and time consuming treatments. A relevant example is represented by UTI.

In this work we investigate the temporal evolution of LUTS as recorded at subsequent clinic attendance visits. We do so accounting for covariates and for correlations among symptoms. In particular, we are interested in controlling for the presence of pyuria in the urine, which is the best biomarker for UTI, in order to obtain robust estimates of the parameters which govern the temporal evolution of the symptoms. For this purpose we analyse a dataset recording the presence of at least one symptom in each of the four categories of LUTS via binary indicators in 1015 patients at 4 different attendance visits. Data have been collected at the *Lower Urinary Tract Service Clinic* (Whittington Hospital, London, UK). Furthermore, at each attendance visit indicators for the presence of pyuria have been also recorded together with the age of the patients.

From a statistical point of view, the task above can be framed in terms of modelling correlated binary variables, where the correlation is among symptoms. The problem of modelling correlated binary variables is frequent in applications and a number of different solutions have been proposed in the literature. One of the most common strategies involves the introduction of continuous latent variables, which are related to the binary variables via thresholds. The success of this class of models is given by the possibility of including complex structures in the latent variables whose distributions are chosen to facilitate posterior inference. Examples are represented by the Probit and Logit models (Albert and Chib, 1993), for which multivariate extension are available (Ashford and Sowden, 1970; Chib and Greenberg, 1998; O'brien and Dunson, 2004; Chen, 2004). In this chapter we focus on the Probit model, which involves Gaussian latent variables discretised at 0: positive and negative values of the latent variables correspond to 1 and 0 at binary level, respectively.

When covariates are available, they can be accommodated in the mean of the latent variables using simple linear regression, which facilitates the interpretation of covariate effect. This simple structure can be generalised also when the latent variable is multivariate, *i.e.* when the objective is to model vectors of binary variables. This can be achieved

by imposing a *seemingly unrelated regressions* (SUR, Zellner, 1962) structure on the latent variables. SUR involves a set of univariate regression models with distinct parameters, but correlated error structure. Once again, time dependence can be incorporated in the model through autoregressive components within the latent variable distributions. An example for univariate binary time series is the work by Giardina et al. (2011).

Recently, assumptions on the latent variables distribution have been relaxed by introducing nonparametric distributions (see Jara et al., 2007). These often involve *Dirichlet process mixture* (DPM) models (Lo, 1984). DPM model consists in convoluting a kernel density indexed by some parameter with a *Dirichlet process* (DP; Ferguson, 1973, Antoniak, 1974; Sethuraman, 1991) distributed discrete random measure. DPM of latent distributions have been employed to model vectors of binary variables (Jara et al., 2007) and univariate binary time series (Di Lucca et al., 2013). In this setting covariates can be included as fixed regression effects (as shown in Jara et al., 2007) in all mixture components favouring the interpretability of the regression coefficients or within the weights of the mixture components modelling jointly the latent variables and the covariates (see DeYoreo et al., 2015). Other related solutions can be found for multicategorical discrete outcomes, *e.g.* Kottas et al. (2005) and DeYoreo and Kottas (2017a). A relevant contribution for this work is in DeYoreo and Kottas (2017b), where a model for time series of univariate ordinal categorical variables is proposed including covariate information through an appropriate model and capturing the time evolution via a time-evolving version of the *dependent Dirichlet process* (DDP, MacEachern, 1999 and 2000).

In this work we propose a model for LUTS which employs latent variables, whose distributions are assumed to be semi-parametric. Risk factors for LUTS and a lagged latent components are included in the latent variables distribution via a SUR model, where the different levels of SUR are indexed to the different symptoms. Intercepts of the regressions and the covariance matrix included in the joint distribution of the error terms are assumed to be component-specific in an infinite mixture model where the mixture weights vary (and are correlated) across different pyuria states and for first visits and follow-ups. This produces the effect of a non-linear regression of the pyuria and visit indicators on the mean and covariance matrix of the latent variables. In this way we are able to achieve more robust estimates for the other linear regression coefficients included in the model, especially for the autoregressive coefficients, in a fashion similar to what has been described by Papageorgiou et al. (2015). The results of the data analysis highlight the different behaviours of the symptoms, both in terms of correlations with other symptoms and in terms of recurrence of symptoms. A relevant connection has been found between the presence of pyuria and the probability of observing voiding symptoms, which in turn is highly correlated with stress incontinence and pain symptoms. In addition, correlations between pairs of symptoms do not seem constant for all patients and in some case they are affected by the presence of pyuria.

The rest of the work is organised as follows. In Section 2 we present the detail of the proposed model, while in Section 3 we describe a *Markov chain Monte Carlo* (MCMC) algorithm for sampling from the posterior distributions of the parameters. The results of the application on the LUTS data are shown in Section 4. We conclude with a discussion in Section 5.

2 A semi-parametric model for binary variables

We consider a study involving N patients, for which a $D-$dimensional binary vector is recorded at T subsequent time points. We denote the collection of all binary records with $\boldsymbol{Y}$, an array having dimensions $N \times D \times T$. Let $y_{i,d,t}$ and $\boldsymbol{y}_{i,t}$ denote a single entry and a row of binary variables of $\boldsymbol{Y}$ corresponding to patient i at time t, respectively.

2.1 Dynamic multivariate Probit model

We model $\boldsymbol{Y}$ introducing an array of correlated continuous latent variables, which we denote with $\boldsymbol{Z}$, having entries $z_{i,d,t} \in \mathbb{R}$ and such that:

$$y_{i,d,t} = 1 \text{ if and only if } z_{i,d,t} \geq 0.$$

Given the condition above, we can write the likelihood of $\boldsymbol{Y}$ as

$$\Pr(\boldsymbol{Y} \mid \boldsymbol{\Theta}) = \int \prod_{\{y_{i,d,t}=1\}} \mathbb{I}_{[0,+\infty)}(z_{i,d,t}) \prod_{\{y_{i,d,t}=0\}} \mathbb{I}_{(-\infty,0)}(z_{i,d,t}) F(\mathrm{d}\boldsymbol{Z} \mid \boldsymbol{\Theta}),$$

where $\mathbb{I}_A(z)$ is an indicator function taking value equal to 1 if z is contained in A and $F(\boldsymbol{Z} \mid \boldsymbol{\Theta})$ is the joint distribution of $\boldsymbol{Z}$, parameterised by $\boldsymbol{\Theta}$.

We recall that T represents the time dimension of $\boldsymbol{Y}$ and we assume for the density of $F(\boldsymbol{Z} \mid \boldsymbol{\Theta})$, which we denote by $f(\boldsymbol{Z} \mid \boldsymbol{\Theta})$, the following factorisation

$$f(\boldsymbol{Z} \mid \boldsymbol{\Theta}) = \prod_{i=1}^{N} \left\{ f(\boldsymbol{z}_{i,1} \mid \Theta_{i,1}) \prod_{t=2}^{T} f(\boldsymbol{z}_{i,t} \mid \boldsymbol{z}_{i,t-1}, \Theta_{i,t}) \right\}, \tag{2.1}$$

where $\boldsymbol{z}_{i,t} = (z_{i,1,t}, \ldots, z_{i,D,t})$, *i.e.* the row of $\boldsymbol{Z}$ corresponding to the $i-$th patient at time t. The latter equation imposes a Markov structure to the distribution of the latent variables, which directly determines the distribution of $\boldsymbol{Y}$.

A computationally convenient assumption is to assume $f(\cdot)$ to be a multivariate Normal distribution. This has a number of implications. First, for any given time t, the probability of observing $\boldsymbol{y}_{i,t}$ can be calculated as an integral under a multivariate Normal distribution, as in multivariate Probit models. Secondly, the Markov structure assumed in (2.1) can be easily accommodated within the model using appropriate autoregressive

terms. Furthermore, covariates can be easily included in the model. Finally, the Normal assumption simplifies the calculations, allowing the use of standard algorithms, such as the one proposed by Albert and Chib (1993), for posterior inference.

Let $\boldsymbol{X}$ denote an array of dimension $N \times P \times T$, containing records of P time-dependent covariates. We write $\boldsymbol{x}_{i,t}$ to indicate the row of $\boldsymbol{X}$ corresponding to the $i-$th patient at time t. Recalling (2.1), we specify the following distribution for the latent variable at time 1:

$$f(\boldsymbol{z}_{i,1} \mid \Theta_{i,1} = (\boldsymbol{\alpha}_{i,1}, \Lambda, \Sigma_{i,1})) = \mathrm{N}_D(\boldsymbol{z}_{i,1} \mid \boldsymbol{\alpha}_{i,1} + \Lambda \boldsymbol{x}'_{i,1}, \Sigma_{i,1}), \qquad (2.2)$$

where $N_D(\cdot)$ is the $D-$dimensional Normal distribution, $\boldsymbol{\alpha}_{i,1}$ is a vector of intercepts of length D and Λ is a $D \times P$ matrix of regression coefficients. Similarly, we specify the following transition density for $t = 2, \ldots, T$:

$$f(\boldsymbol{z}_{i,t} \mid \boldsymbol{z}_{i,t-1}, \Theta_{i,t} = (\boldsymbol{\alpha}_{i,t}, \Lambda, \Gamma, \Sigma_{i,t})) = \mathrm{N}_D(\boldsymbol{z}_{i,t} \mid \boldsymbol{\alpha}_{i,t} + \Lambda \boldsymbol{x}'_{i,t} + \Gamma \boldsymbol{z}'_{i,t-1}, \Sigma_{i,t}), \qquad (2.3)$$

where Γ is a $D \times D$ matrix containing the autoregressive coefficients.

The model described above is connected to the SUR model in that at a time point, given all parameters and lagged latent components, the distributions above imply D distinct regression on the means of the latent variables. The latter are linked together via the error distributions governed by the covariance matrix $\Sigma_{i,t}$.

Λ and Γ are assumed to be constant across time. This simplifies the interpretability of the coefficients. In particular, simple Bayesian hypothesis testing on temporal dependence among the binary variables can be performed and this is an important requirement for our motivating application. However, extensions including temporal dependent versions of the matrices of coefficients Λ_t and Γ_t can be easily specified.

2.2 Nonparametric prior model

The quantities $\boldsymbol{\alpha}_{i,t}$ and $\Sigma_{i,t}$ characterise the baseline probability of observing $\boldsymbol{y}_{i,t}$. We want to employ a prior distribution which is flexible enough to capture possible heterogeneity across patients at different time point.

Let us introduce an additional array of covariates, $\boldsymbol{U}$. This is an $N \times R \times T$ array, where $\boldsymbol{u}_{i,t}$ is a row of $\boldsymbol{U}$ which encodes the information about the $i-$th patient at time t, such as treatment arm, risk group and other common indicators which may evolve overtime. This is not uncommon in biostatistics when clinicians classify patients into different classes of risk for developing a specific disease based on clinical history and individual characteristics.

We include via a flexible model the information contained in $\boldsymbol{U}$ in the baseline probability of $\boldsymbol{y}_{i,t}$. Using an approach similar to the one employed in mixed-effect modelling,

we assume each group defined by $\boldsymbol{U}$ to have a specific random prior distribution as follows

$$(\boldsymbol{\alpha}_{i,t}, \Sigma_{i,t}) \mid G_{\boldsymbol{u}_{i,t}} \sim G_{\boldsymbol{u}_{i,t}},$$

where $G_{\boldsymbol{u}_{i,t}}$ is a discrete distribution of the form

$$G_{\boldsymbol{u}_{i,t}} = \sum_{h=1}^{\infty} w_{h,\boldsymbol{u}_{i,t}} \delta_{\boldsymbol{\alpha}_h^*, \Sigma_h^*}, \tag{2.4}$$

where, for each $\boldsymbol{u}_{i,t}$, $\sum_h w_{h,\boldsymbol{u}_{i,t}} = 1$, and δ_a denotes a unit point mass at a. The use of such random prior distribution implies that the latent variable density can be written as

$$
\begin{aligned}
f(\boldsymbol{z}_{i,t} \mid \boldsymbol{z}_{i,t-1}, \Lambda, \Gamma, G_{\boldsymbol{u}_{i,t}}) &= \int f(\boldsymbol{z}_{i,t} \mid \boldsymbol{z}_{i,t-1}, \Lambda, \Gamma, \boldsymbol{\alpha}_{i,t}, \Sigma_{i,t}) \mathrm{d}G_{\boldsymbol{u}_{i,t}}(\boldsymbol{\alpha}_{i,t}, \Sigma_{i,t}) \\
&= \sum_{h=1}^{\infty} w_{h,\boldsymbol{u}_{i,t}} N_D(\boldsymbol{z}_{i,t} \mid \boldsymbol{\alpha}_h^* + \Lambda \boldsymbol{x}_{i,t}' + \Gamma \boldsymbol{z}_{i,t-1}', \Sigma_h^*),
\end{aligned}
$$

which is an infinite location-scale mixture of multivariate Normal distributions, having weights which vary according to the components in $\boldsymbol{u}_{i,t}$.

Specifying a prior of $G_{\boldsymbol{u}_{i,t}}$ is equivalent to finding suitable prior distributions for the collection of the weights, $w_{h,\boldsymbol{u}_{i,t}}$, and for the locations, $\boldsymbol{\alpha}_h^*$ and Σ_h^*.

Starting from the weights, we need a stochastic process prior indexed at various levels of $\boldsymbol{u}_{i,t}$, whose realisations are distributions over an infinite dimensional simplex. Furthermore, it is desirable to borrow strength across the different groups of observations implied by $\boldsymbol{U}$. Barcella et al. (2016) introduce the *dependent generalised Dirichlet process* (DGDP), a process over collections of distributions which can be considered an extension of the DDP. We opt for using the same idea of the DGDP also here, but we will mention alternative solutions in Section 5. DGDP assumes a particular stick-breaking process prior for the weights of (2.4) where $w_{1,\boldsymbol{u}_{i,t}} = v_{1,\boldsymbol{u}_{i,t}}$ and

$$w_{h,\boldsymbol{u}_{i,t}} = v_{h,\boldsymbol{u}_{i,t}} \prod_{l<h} (1 - v_{l,\boldsymbol{u}_{i,t}}), \text{ for } h = 2, 3, \dots$$

and

$$v_{1,\boldsymbol{u}_{i,t}}, v_{2,\boldsymbol{u}_{i,t}}, \dots \mid \boldsymbol{\mu}, \phi \sim \text{Beta}(v_{h,\boldsymbol{u}_{i,t}} \mid \phi\mu(\boldsymbol{u}_{i,t}), \phi(1 - \mu(\boldsymbol{u}_{i,t}))),$$

for all possible values of $\boldsymbol{u}_{i,t}$. In the latter equation, ϕ is a positive parameter and $\mu(\cdot)$ is random mean function which maps into the interval $(0,1)$. Here we assume that $\mu(\boldsymbol{u}_{i,t}) = \text{logit}^{-1}(\boldsymbol{\mu}\boldsymbol{u}_{i,t})$, where $\boldsymbol{\mu}$ is a vector of unknown real parameters, but alternative mean functions can be included. The almost sure discreteness of DGDP realisations imposes a clustering structure of the attendance visits: observations sharing the same value of $\boldsymbol{\alpha}_h^*$ and Σ_h^* can be interpreted as a cluster. Compared to DDP (which will

in principle produce an equivalent clustering of attendance visits), DGDP allows extra flexibility in the distribution of the clusters especially in terms of number and size of clusters, due to the richer parameterisation of the DGDP compared to the traditional DDP.

Independently from the generation of the weights, the locations, which are shared among the $G_{\boldsymbol{u}_{i,t}}$ for different values of $\boldsymbol{u}_{i,t}$, are generated as follows. We first assume

$$\boldsymbol{\alpha}_h^* \sim \mathrm{N}_D(\boldsymbol{m}_D, \sigma_\alpha^2 I_D), \text{ for } h = 1, 2, \ldots$$

while Σ_h^* requires likelihood identifiability conditions imposed by the thresholding of the latent variables. In order to avoid over-restrictive constraints on the covariance matrix, we follow the works by Jara et al. (2007) and Pourahmadi (1999) where the conditional variances are constrained. We write $\Sigma_h^{*-1} = L_h^{*'} I_D L_h^*$, where L_h^* is a lower triangular matrix, with ones on the diagonal and unconstrained values on the non-zero entries. Let ν_h^* be the vector of free parameters in L_h^*, such that ν_h^* has $D_\nu = D(D-1)/2$ components, we assume

$$\nu_h^* \sim \mathrm{N}_{D_\nu}(\boldsymbol{m}_{D_\nu}, \sigma_\nu^2 I_{D_\nu}), \text{ for } h = 1, 2, \ldots$$

In this work we assume shared locations for all $\boldsymbol{u}_{i,t}$ following the argument discussed in Hatjispyros et al. (2015), who show that when constructing dependent random measures this construction with varying weights can lead to random measures with arbitrary degree of proximity. However, a possible extension can include covariate dependent locations, in this case requiring the use of suitable stochastic process priors indexed by the covariates.

2.3 Prior distribution specification

The model described above requires the specification of prior (and hyperprior) distributions for the remaining unknown parameters. Let $\lambda_{d,p}$ and $\gamma_{d,d'}$ denote a single entry in matrix Λ and Γ, respectively. We assume for d and $d' = 1, \ldots, D$ and $p = 1, \ldots, P$

$$\lambda_{d,p} \overset{iid}{\sim} \mathrm{N}(\lambda \mid m_\lambda, \sigma_\lambda^2)$$

$$\gamma_{d,d'} \overset{iid}{\sim} \mathrm{N}(\gamma \mid m_\gamma, \sigma_\gamma^2).$$

The hyperprior distributions for the values of $\boldsymbol{\mu}$ and ϕ are set as follows

$$\boldsymbol{\mu} \sim \mathrm{N}_R(\boldsymbol{\mu} \mid m_\mu, \sigma_\mu^2 I_R)$$

$$\phi \sim \mathrm{Gamma}(\phi \mid a_\phi, b_\phi),$$

where the latter is a Gamma distribution with expectation a_ϕ/b_ϕ. The choice of these prior distributions is mainly motivated by computational reasons.

3   Posterior inference

Posterior inference can be performed using MCMC methods. For the model described in Section 2, samples from posterior distribution of the parameters can be approximated using a Metropolis-within-Gibbs algorithm, using a finite approximation of the nonparametric prior distribution in (2.4) up to H atoms. This strategy is usually referred to as *blocked Gibbs sampler* (Ishwaran and James, 2001). A summary of the main steps of the algorithm can be found in Appendix A. Different alternatives have been discussed in the literature to determine a value of H when the random distribution follows a DP or more general processes. See, for example, Ohlssen et al. (2007) and Ishwaran and Zarepour (2000). A simple strategy that can be adapted to the case of DGDP consists in checking that the expectation and variance of the weight $w_{H,\boldsymbol{u}}$ are adequate for the values of $\boldsymbol{\mu}$ and ϕ explored by the MCMC. Truncating the stick-breaking prior to a finite number of components H reduces model complexities and simplifies the construction of the algorithm for posterior inference, also allowing the use of softwares for Bayesian inference such as `JAGS`. An alternative approach for posterior sampling which does not involve an approximation of the nonparametric prior is based on the slice sampling algorithm described in Kalli et al. (2011) and Walker (2007).

4   Application: Lower Urinary Tract Symptoms

In this section we present background information, exploratory analysis and results of the application of the model described in the previous sections to the study of the evolution of LUTS.

4.1   Dataset

The dataset employed in the analysis contains information about 1015 female patients affected by LUTS, who have attended at four subsequent visits the Lower Urinary Tract Service Clinic (Whittington Hospital, London, UK). At each attendance visit the following information is collected: the date of the visit, the age of the patient, the presence of urgency symptoms (binary), presence of voiding symptoms (binary), presence of pain symptoms (binary), presence of stress incontinence symptoms (binary), and the count of *white blood cells* (WBC) in a sample of urine.

The most frequently observed symptom is urgency, which affects 72.32% of the patients at the first attendance visit and in 61.77% at follow-up visits. The least frequent symptom is stress incontinence, which is observed in 37.14% of first visits and in 34.93% of follow-ups. Pain and voiding symptoms are observed in 54.68% and 40.00% of the first visits and 44.83% and 34.93% of the follow-ups, respectively.

WBC counts are used to assess the presence of pyuria, which is considered the best biomarker for UTI. While evidence has been collected that relates a WBC count larger than or equal to one to the presence of infection, the threshold of ten WBC in a urine sample is the sensitivity of common dipstick tests used to assess the presence of pyuria. Following these thresholds we generate two binary indicators from WBC and we refer to them as *mild pyuria* in the case of $1 \leq \text{WBC} \leq 9$, and to *severe pyuria* for $\text{WBC} \geq 10$. These indicators can be interpreted in terms of severity of the infection: in fact, large WBC counts indicate a high degree of inflammation, and thus may lead to complications. Mild pyuria has been observed in 17.86% of the attendance visits, while severe pyuria has been recorded in 18.74% of the cases. Considering the first attendance visits exclusively, these become 18.62% and 21.97%, respectively.

The average number of days between attendance visits is approximately 91 days (standard deviation equal to 143.22), where the shortest period is observed between the first two visits with an average of 85 days (standard deviation equal to 147.58) and the longest one is recorded between the last two visits with an average of 95 days (standard deviation equal to 135.07). Finally, the age of the patients at first attendance visits is on average 54, with a sample standard deviation of 17.27.

In terms of treatment regimes, all patients have been treated with a combination of antimuscarinic and bladder retaining after the first attendance visits until the fourth one. Furthermore, patients diagnosed with mild and severe pyuria have been treated with antibiotics.

4.2   Data pre-processing and choice of hyperparameters

We denote the array containing information on the presence of the symptoms as $\boldsymbol{Y}$, which has dimension $1015 \times 4 \times 4$ (corresponding to the number of patients, the number of symptoms and the number of attendance visits, respectively), and contains binary observations. The symptoms recorded in a row $\boldsymbol{y}_{i,t}$ of $\boldsymbol{Y}$ are abbreviated as U, P, S, and V, standing for urgency, pain, stress incontinence and voiding symptoms, respectively. Consequently, in the following paragraphs the indexes d and d' take value in $\{U, P, S, V\}$.

We denote with $\boldsymbol{X}$ the array containing information about the age of the patients and the time in days from the first attendance visits for all subsequent visits. In this application we only consider the age at the first attendance visits instead of the age when each attendance visit takes place. Therefore, $\boldsymbol{X}$ has dimension $1015 \times 2 \times 4$ corresponding to the number of patients, the number of covariates and the number of attendance visits. Across attendance visits, the first covariate (*Age*) remains constant, while the second one (*Days*) has entries equal to zero for the first attendance visits. For each attendance visit, the two covariates have been centred and rescaled to have mean and variance equal to 0 and 1, respectively.

We include as covariates in $\boldsymbol{U}$ the indicators for mild and severe pyuria as well as an indicator for first attendance visits and intercept terms. $\boldsymbol{U}$ is then an array with dimension $1015 \times 4 \times 4$. Consequently, considering the row vectors $\boldsymbol{u}_{i,t}$ of $\boldsymbol{U}$ and all possible values they can take, we can identify six groups of attendance visits which are all observed in the data set.

Both $\boldsymbol{X}$ and $\boldsymbol{U}$ are used as covariates, however their effect on the latent variables are different. In particular, the effect of $\boldsymbol{U}$ is non-linear via the intercepts $\boldsymbol{\alpha}$ and the covariance matrix $\boldsymbol{\Sigma}$. On the other hand, the entries of $\boldsymbol{X}$, similarly to the autoregressive terms, affect the mean of latent variables linearly. The decision of which covariates to assign to $\boldsymbol{U}$ and to $\boldsymbol{X}$ is driven by the specific application. In particular, $\boldsymbol{U}$ captures the natural clustering of the observations based on the levels of the UTI and first attendance visits and follow-ups.

We set the following hyperparameter values: $\boldsymbol{m}_\alpha = \boldsymbol{0}_D$ (where $\boldsymbol{0}_a$ denotes a vector with a components all equal to zero); $\boldsymbol{m}_\nu = \boldsymbol{0}_{D_\nu}$; $m_\lambda = m_\gamma = m_\mu = 0$; $\sigma_\alpha^2 = \sigma_\nu^2 = \sigma_\lambda^2 = \sigma_\gamma^2 = 100$; $\sigma_\mu^2 = 1$; $a_\phi = b_\phi = 1$. We use the algorithm described in Section 3 for posterior inference. This requires the specification of a truncation level H for the DGDP and we set it equal to 20. We initialise the algorithm drawing random values from the prior distribution of each parameter and we run the algorithm for $50\,000$ iterations, after a burning period of $10\,000$, and we save every tenth sample. Convergence of MCMC has been assessed using trace plots and autocorrelation.

4.3 Results

Before discussing the results of the application, we notice that the marginal probability of observing a certain symptom is equal to

$$\Pr(y_{i,d,t} = 1 \mid \alpha_{i,d,t}, \boldsymbol{\lambda}_d, \boldsymbol{\gamma}_d, \sigma_{i,d,t}, \boldsymbol{z}_{i,t-1}) = \Phi\left(\frac{\alpha_{i,d,t} + \boldsymbol{\lambda}_d \boldsymbol{x}'_{i,t} + \boldsymbol{\gamma}_d \boldsymbol{z}'_{i,t-1}}{\sigma_{i,d,t}}\right),$$

where $\boldsymbol{\lambda}_d$ and $\boldsymbol{\gamma}_d$ are rows of Λ and Γ, respectively. This shows that $\alpha_{i,d,t}$ controls the baseline probability of the symptom d at visit t, together with $\sigma_{i,d,t}$, the square root of the d–th diagonal component of $\Sigma_{i,t}$. When the covariates $\boldsymbol{x}_{i,t}$ and $\boldsymbol{z}_{i,t}$ have components equal to zero, the sign of $\alpha_{i,d,t}$ determines if the probability of observing the symptoms is larger than 0.5, which happens when the parameter is positive. The dependence among symptoms can be evaluated computing the conditional distributions of components of the latent vector, *i.e.* $p(z_{i,d,t} \mid \boldsymbol{z}_{i,-d,t}, \ldots)$, where $\boldsymbol{z}_{i,-d,t}$ is obtained removing $z_{i,d,t}$ from $\boldsymbol{z}_{i,d,t}$.

We discuss posterior inference for the parameters of the proposed model distinguishing between the individual effects, $\boldsymbol{\alpha}_{i,t}$ and $\Sigma_{i,t}$, and the shared effects (*i.e.* shared among all patients), Λ and Γ.

Individual effects

We first consider the posterior distribution of the vector $\boldsymbol{\alpha}_{i,t}$. The discrete prior distribution induces ties among the values $\boldsymbol{\alpha}_{i,t}$ of different patients. This implies that we can cluster patients at time t based on the unique values $\boldsymbol{\alpha}_h^*$ and Σ_h^*, i.e. two patients sharing the same value of $\boldsymbol{\alpha}_h^*$ and Σ_h^* belong to the same cluster. This is a crucial aspect of the proposed modelling strategy both in terms of the ability to capture the heterogeneity in the data and the interpretability of the results as the clustering of the observations can provide further insight into the phenomenon under investigation.
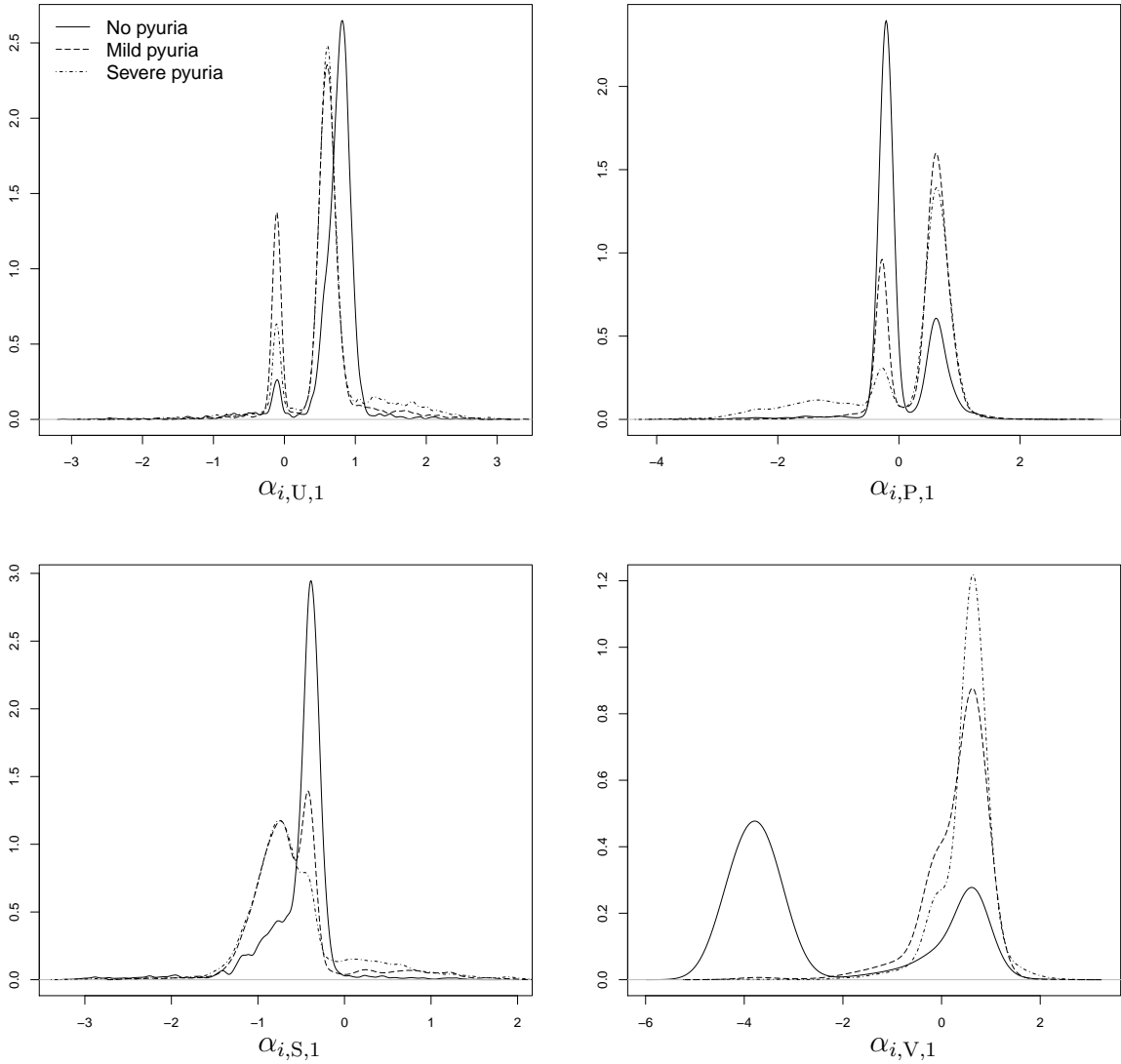

Figure 1: Posterior density of $\alpha_{i,d,t}$, $d = \{U, P, S, V\}$, when t is the first attendance visit, for different pyuria levels.

Figure 1 shows the posterior distribution of the intercepts for the patients at first attendance visits for different levels of pyuria. The difference between the posterior densities for the patients without pyuria and those with mild or severe pyuria is evident.

11

On the other hand, densities corresponding to patients showing mild and severe pyuria are similar. In particular, the presence of pyuria (either mild or severe) leads to higher posterior mean for the latent variables associated with pain and voiding symptoms. The opposite happens for urgency and stress incontinence symptoms, for which the presence of pyuria reduces the mean of the associated latent variables. The largest effect of pyuria is observed for voiding symptoms.
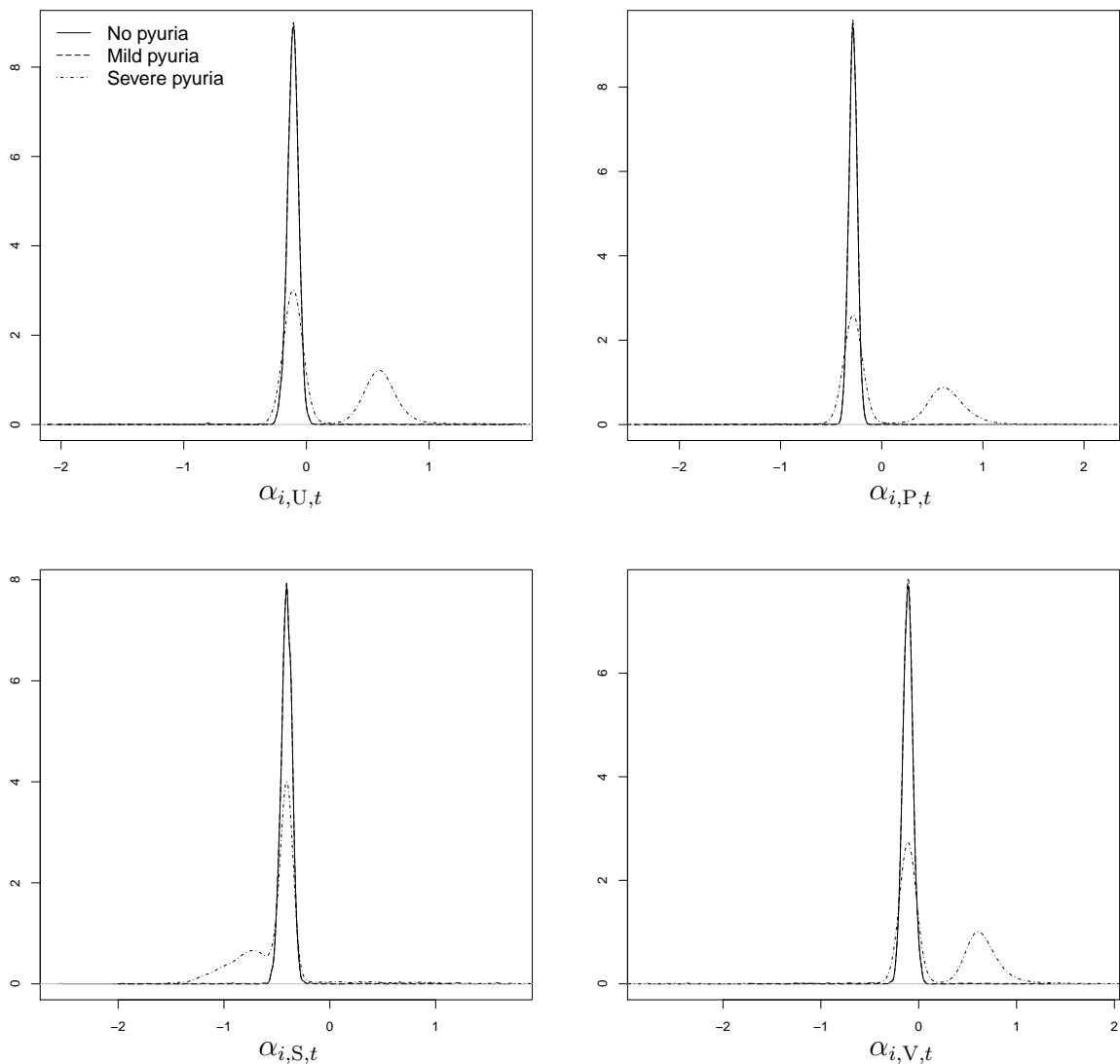

Figure 2: Posterior density of $\alpha_{i,d,t}$ with $d = \{U, P, S, V\}$ and $t = 2, 3, 4$, for different pyuria levels.

For all attendance visits after the first one the posterior densities are presented in Figure 2. Recall that patients after the first attendance visit have all been treated for LUTS. In this case the posterior expectations of $\alpha_{i,d,t}$'s are all negative for the cases with no and mild pyuria which show almost identical posterior distributions. Differently, the presence of severe pyuria strongly affects the posterior densities of $\alpha_{i,d,t}$: positive marginal

posterior expectations are observed for urgency, pain and voiding symptoms.

We now discuss the results for the parameters representing the the entries of the covariance matrix, *i.e.* $\Sigma_{i,t}$, which are also subject-specific. Our focus on the posterior distribution of the correlations' coefficients, which can be easily computed from $\Sigma_{i,t}$. Similarly to $\boldsymbol{\alpha}_{i,t}$, the posterior distributions of these coefficients are indexed by the values of $\boldsymbol{u}_{i,t}$. Given that we are including in our model four symptoms, we have six pairwise correlations, which vary across different pyuria levels and between first attendance visits and follow-ups. The posterior density estimates for each correlation are reported in Figure 3 which shows the correlations estimated at first attendance visits and follow-up visits for different pyuria levels. On the y-axis, we show correlations between different pairs of symptoms, which are denoted by their initial letter.
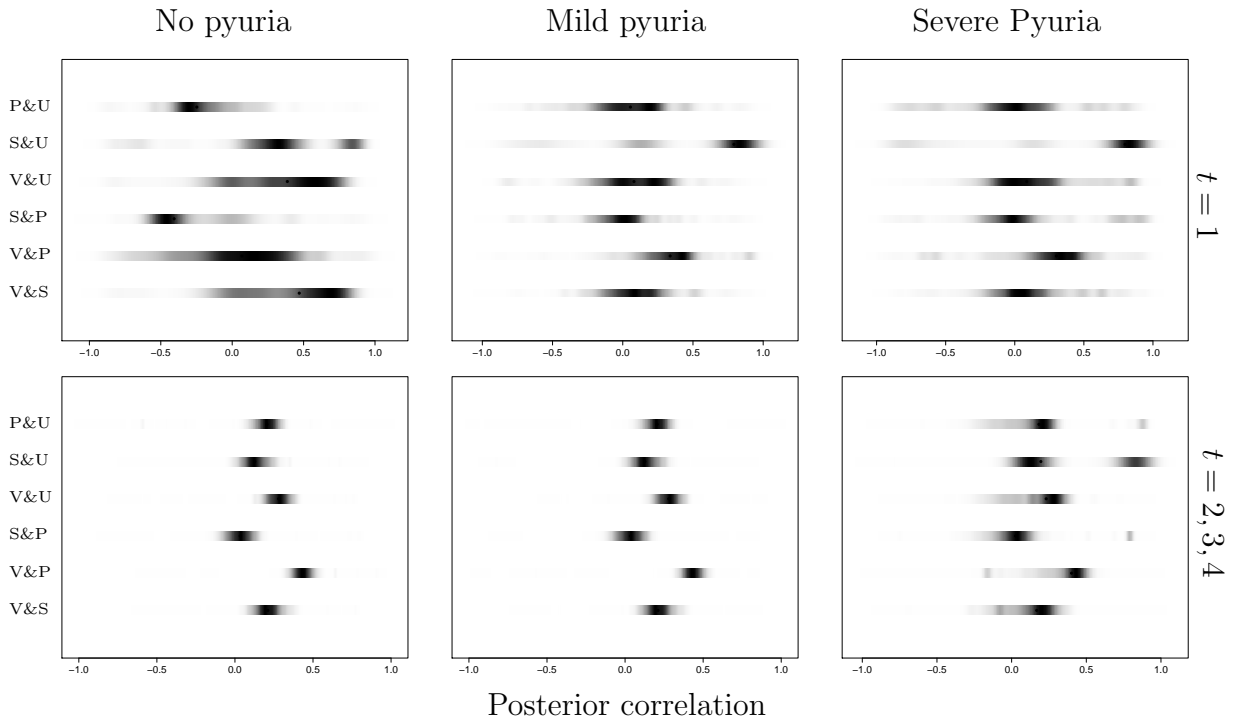

Figure 3: Posterior densities of the correlation between different pairs of latent variables (corresponding to different symptoms) at first attendance visit (top panels) and at follow-up visits (bottom panels), for no pyuria (left panels), mild pyuria (central panels) and severe pyuria (right panels).

The covariate that has the strongest impact on these posterior distributions is the indicator of first attendance visit. This can be seen by comparing the densities associated to each pair of symptoms in Figure 3, where the top panels correspond to the first visit attendance and the bottom ones to follow-ups. Obviously, part of the difference may be due to the fact that correlations for follow-up visits are estimated using a larger number of observations compared to those at first attendance visits. In some cases the posterior expectation of the correlations move toward zero after the first visit attendance (S & U

and S & P). In the other cases, we notice a positive increase of the correlation values (P & U and V & P).

In order to show the advantages of the proposed method in terms of estimating the correlations among symptoms, we compare the results above with those obtained fitting a multivariate version of the parametric Probit model. In its traditional formulation, a Probit model allows covariates to affect the probability of the binary outcomes only through a regression model placed on the mean of the latent variable distribution, whereas the covariance matrix governing the correlations among the latent variables is assumed to be constant. Given the latter assumption, this model cannot detect different levels of correlations among the binary variables for different covariate levels.
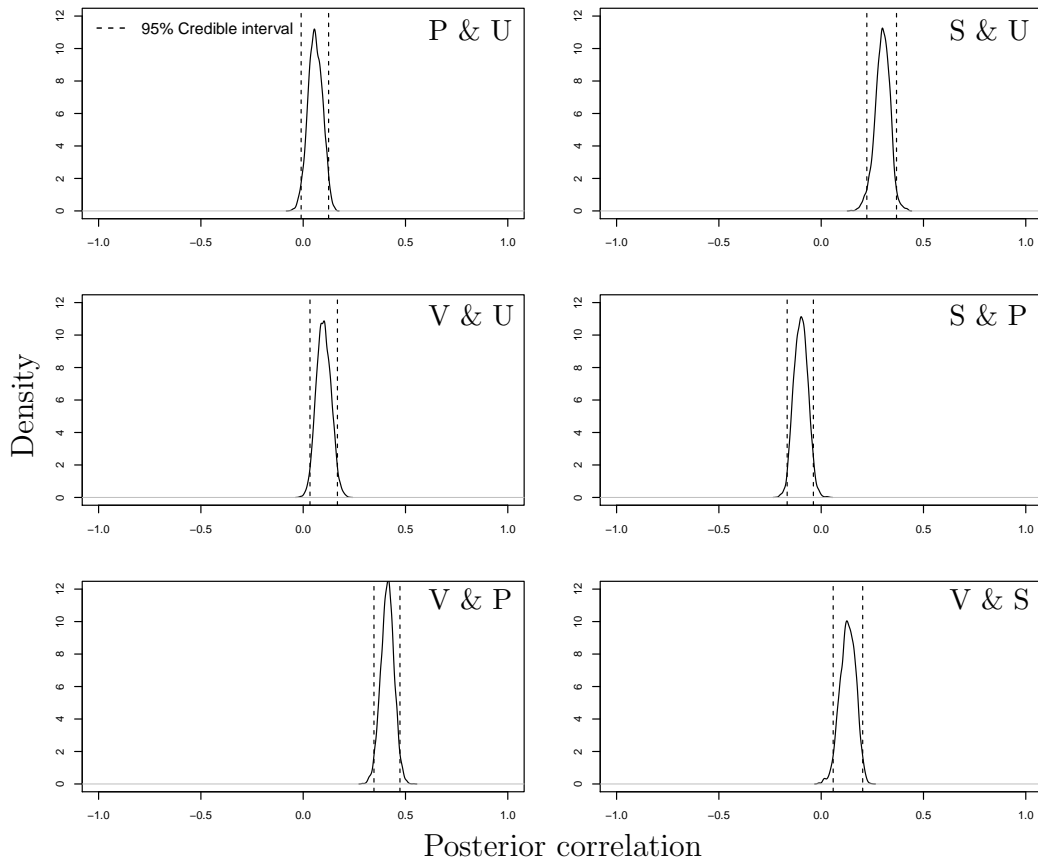

Figure 4: Posterior densities of the correlation between latent variables, estimated using a parametric multivariate Probit model.

In Figure 4, we show the posterior distributions of the correlation between the latent variables using a parametric Probit model. We focus on the pairs of latent variables corresponding to S & P, P & U and S & U. For the first pair of symptoms the estimated posterior distribution is located around -0.1, with a 95% credible interval that does not contain 0. This finding is in contrast to what can be seen in Figure 3, where negative correlation around -0.5 is displayed for the combination of covariates corresponding to

14

the first visit attendance and no pyuria, while for all other levels of pyuria, independently from the visit number, the posterior distribution of the correlation is around 0. Similarly, the posterior distribution of the correlation of P & U in Figure 4 has 95% credible interval containing 0 while the corresponding distributions in the nonparametric model are centred at either negative values for no pyuria at first attendance visit or positive values for all levels of pyuria at follow-up visits. Moreover, both the parametric and nonparametric model estimate posterior distributions for the correlation of S & U concentrated on positive values, but the nonparametric model better accounts also for heterogeneity in the patient population estimating bimodal posterior distributions (see for example the results for severe pyuria into follow-up visits). This comparison highlights the ability of the proposed model to capture different patterns in the correlations of the latent variables which depend on covariate levels, leading to a deeper understanding of the problem and often to different conclusions from the parametric Probit model.

Shared effects

We now discuss the estimated posterior distributions of the parameters shared by all patients.

Table 1: Summary of the posterior distributions of $\gamma_{d,d'}$, with d referring to the latent dependent variables of the d-th symptoms and d' to the latent independent variables of the d'-th symptoms. Effective Sample Size (ESS) is reported in the last column.

	mean	sd	2.5%	25%	50%	75%	97.5%	ESS
$\gamma_{U,U}$	0.74	0.04	0.67	0.72	0.74	0.76	0.82	5000
$\gamma_{U,P}$	0.02	0.03	-0.04	-0.00	0.02	0.04	0.08	5431
$\gamma_{U,S}$	0.10	0.02	0.06	0.09	0.10	0.12	0.15	4683
$\gamma_{U,V}$	-0.09	0.02	-0.12	-0.10	-0.09	-0.07	-0.06	5000
$\gamma_{P,U}$	-0.03	0.03	-0.09	-0.05	-0.03	-0.01	0.03	5000
$\gamma_{P,P}$	0.70	0.03	0.65	0.69	0.70	0.72	0.76	6280
$\gamma_{P,S}$	-0.06	0.02	-0.10	-0.08	-0.06	-0.04	-0.01	4216
$\gamma_{P,V}$	0.01	0.01	-0.02	-0.00	0.01	0.02	0.04	5000
$\gamma_{S,U}$	-0.03	0.03	-0.10	-0.05	-0.03	-0.01	0.04	5000
$\gamma_{S,P}$	0.03	0.04	-0.04	0.00	0.03	0.05	0.10	4925
$\gamma_{S,S}$	0.84	0.04	0.78	0.82	0.84	0.87	0.93	5000
$\gamma_{S,V}$	0.05	0.02	0.01	0.03	0.05	0.06	0.08	5000
$\gamma_{V,U}$	-0.13	0.04	-0.23	-0.16	-0.13	-0.10	-0.06	4673
$\gamma_{V,P}$	-0.01	0.05	-0.10	-0.05	-0.02	0.01	0.09	5000
$\gamma_{V,S}$	0.05	0.03	-0.01	0.03	0.05	0.06	0.10	5229
$\gamma_{V,V}$	0.78	0.03	0.71	0.76	0.78	0.80	0.83	5000

In Table 1, posterior summaries for each entry of the matrix Γ have been reported. These parameters capture the autoregressive effects of the latent variables for each symp-

tom and are assumed to be time-invariant. We also assume that the latent variable of each symptom at a specific attendance visit is affected by the values of the latent variables of all symptoms at the previous attendance visit. In this way we want to control for the temporal interaction among symptoms.

We begin considering the autoregressive effects, *i.e.* the parameters governing the dependence between the latent variables at subsequent time points, namely $\gamma_{U,U}, \gamma_{P,P}, \gamma_{S,S}$ and $\gamma_{V,V}$. The posterior distributions of these parameters are concentrated on positive values and the largest posterior expectations are estimated for stress incontinence and voiding. Pain symptoms instead appear to be the least recurrent having the smallest posterior mean. All posterior 95% credible intervals for these parameters do not contain zero.

On the contrary, cross-effects, *i.e.* the parameters governing the dependence between different symptoms at subsequent attendance visits, are often centred around zero (considering posterior 95% credible intervals). Exceptions are $\gamma_{U,V}$, $\gamma_{V,U}$, $\gamma_{P,S}$ which show a negative effect. Instead, $\gamma_{U,S}$ and $\gamma_{S,V}$ show positive effect.

Table 2: Summary of the posterior distributions of $\lambda_{d,m}$, with $d = \{U, P, S, V\}$ referring to the symptoms and $m = \{Age, Days\}$ to different covariates. Effective Sample Size (ESS) is reported in the last column.

	mean	sd	2.5%	25%	50%	75%	97.5%	ESS
$\lambda_{U,\ Age}$	0.07	0.02	0.03	0.06	0.07	0.09	0.12	5000
$\lambda_{P,\ Age}$	-0.13	0.02	-0.18	-0.15	-0.13	-0.12	-0.09	5000
$\lambda_{S,\ Age}$	0.02	0.03	-0.04	-0.00	0.02	0.04	0.07	5000
$\lambda_{V,\ Age}$	-0.00	0.03	-0.07	-0.02	-0.00	0.02	0.06	5000
$\lambda_{U,\ Days}$	0.06	0.03	0.01	0.04	0.06	0.08	0.12	4410
$\lambda_{P,\ Days}$	0.06	0.02	0.01	0.04	0.06	0.07	0.10	5000
$\lambda_{S,\ Days}$	0.06	0.03	-0.00	0.04	0.06	0.08	0.11	5000
$\lambda_{V,\ Days}$	0.05	0.04	-0.02	0.02	0.05	0.07	0.12	5000

Table 2 reports posterior summaries for the regression coefficients of *Age* and *Days*. The former has a positive effect on the probability of observing urgency symptoms, while negative effect on the probability of observing pain symptoms. Posterior 95% credible intervals for stress incontinence and voiding symptoms are centred around zero.

The period in days from the first visit attendance increases the probability of observing urgency and pain symptoms while it has no evident effect on stress incontinence and voiding symptoms considering 95% credible intervals which contain zero.

Predictive inference

We summarise the main results described in previous sections in Figure 5.

This depicts different trajectories of the marginal probability of observing the different
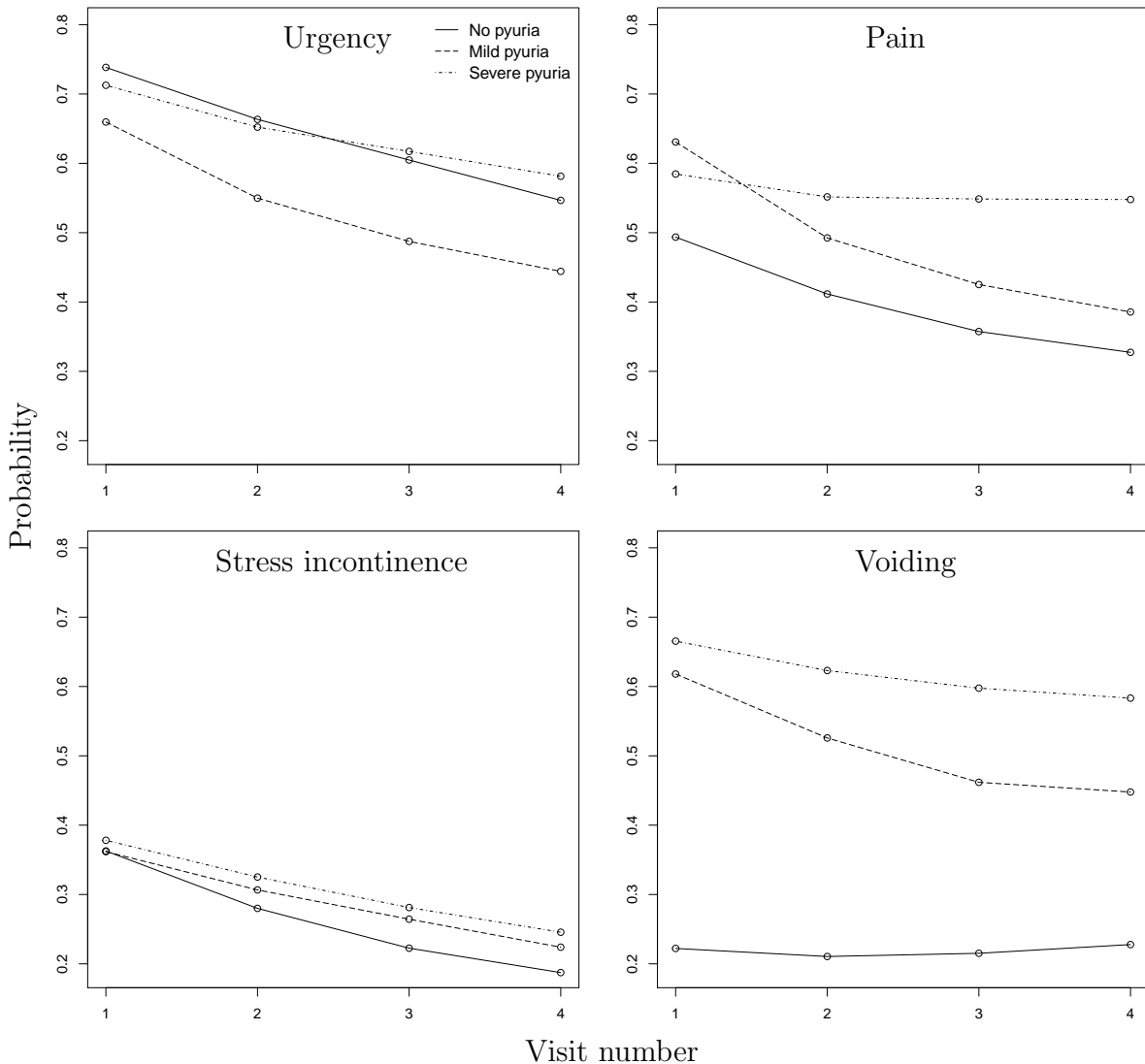
Figure 5: Marginal predictive probability for the four categories of symptoms at four subsequent attendance visits. We consider patients having for all four visits either no pyuria (solid lines) or mild pyuria (dashed lines) or severe pyuria (dotted lines).

categories of symptoms over time, using predictive distributions. We derive numerically the predictive distribution over the trajectories fixing *Age* equal to the mean observed value, as well as the mean value of *Days*. We then explore the probabilities for symptoms for the three groups of patients: no pyuria, mild pyuria and severe pyuria for all four the attendance visits.

The trajectories for all symptoms are decreasing over time, the only exception being voiding symptoms in patients with no pyuria where after a decrease in probability at the second attendance visit, the probability of observing symptoms slightly increase during the last two attendance visits. The group of symptoms which is mostly affected by the presence of pyuria is voiding, followed by pain. Both urgency and stress incontinence symptoms are less affected by the presence of pyuria. In particular, while in general both

mild and severe pyuria increase the probability of observing stress incontinence symptoms, mild pyuria decreases the probability of observing urgency symptoms and severe pyuria initially decreases the probability of observing urgency symptoms compared to the case of no pyuria, but this probability increases over the subsequent attendance visits, relatively to no pyuria patients.

Finally, although the predictive distributions show decreasing trajectories for almost all symptoms, it is hard to conclude on the effectiveness of the treatments administered to the patients. This is because this study include a single treatment arm and it has been designed to collect information only about dependence among symptoms across time and the relation of symptoms with the degree of infection.

5 Discussion

In this work we propose a method for modelling vectors of correlated binary variables evolving over time. The motivation for this work comes from a study of time evolving records of LUTS and connected risk factors. Similarly to a traditional Probit model, we assume the binary variables to be distributed as a discretised version of continuous latent variables. We assume the latent variables to be generated from an infinite mixture model, with weights that vary across a covariate space. Moreover, the time evolution of the symptoms is modelled using autoregressive and cross-effect terms.

This specification induces very flexible distributions for vectors of binary variables while allowing the user to maintain clear interpretation of the parameters of interest. Since the resulting model is a mixture of Gaussian distributions, covariates can be accommodated in the mean of each mixture component as regression terms, as well as autoregressive components.

At the latent variable level, the proposed model can be considered a nonparametric version of a mixed-effect model, where the mixing distributions depends on some covariate level. The choice of the stochastic process prior for the mixing distribution should reflect the type of information we want to include from covariates and the available prior information. In this work we choose the DGDP, for its flexibility and the possibility of including an ANOVA model in the stick-breaking process of the weights. However, when the covariates are continuous the use of DGDP leads to over-parameterised models, suggesting the use of alternative stochastic process priors. Examples are the Probit stick-breaking (Rodriguez and Dunson, 2011), logit stick-breaking (Ren et al., 2011) and kernel stick-breaking (Dunson and Park, 2008).

The results of the data analysis show that the different levels of pyuria strongly affect the probability of observing the symptoms, particularly voiding symptoms which seem to be the most probable group of symptoms to activate in case of infection. The effect of pyuria on the symptoms correlations is instead less clear and varies for different

pairs of symptoms. In order to evaluate the temporal dependence among symptoms we summarise the posterior distribution of the autoregressive coefficients and we find that the most recurrent symptoms appear to be stress incontinence and voiding. We also evaluate cross-effects among symptoms occurring at subsequent times, finding strong relationships between voiding and urgency symptoms and between urgency and stress incontinence symptoms.

The MCMC algorithm for posterior inference in implemented in JAGS through `R` using the the package `R2jags`. For the dataset under analysis and with the parameter settings specified in the paper, the MCMC algorithm takes approximately 15 hours to run on a desktop computer with an *Intel i-7* processor. Alternatively, the algorithm described in Appendix A could be implemented in a low level programming language, leading to considerable gains in computational speed. This is especially advised when larger datasets are involved. Finally, the proposed model formulation has been introduced for dealing with vectors of binary variables observed at different observation times which are assumed to be the same for all individuals, since in our application patients are visited in four occasions. However, the same model formulation can be easily extended to include sequences of binary vectors with heterogenous number of observations across patients.

A   Blocked Gibbs sampler for posterior inference

In this appendix we summarise the main parts of the blocked Gibbs sampler algorithm discussed in Section 3 for posterior inference of the proposed model.

i) *Sample $\boldsymbol{Z}$ given all other parameters.* In the proposed model there are $N \times D \times T$ latent variables, which can be resampled sequentially starting at $t = 1$ using the algorithms for binary Probit models described in Albert and Chib (1993) and Holmes and Held (2006).

ii) *Sample Λ given all other parameters.* These parameters are shared by all $i = 1, \ldots, N$ and $t = 1, \ldots, T$ and can be resampled directly from the full conditionals.

iii) *Sample Γ given all other paramters.* These parameters are shared by all $i = 1, \ldots, N$ and $t = 2, \ldots, T$ and can be resampled directly from the full conditionals.

iv) *Sample $\{(\boldsymbol{\alpha}_{i,t}, \Sigma_{i,t}), \forall i, t\}$ given all other parameters.* This step is based on a finite truncation of the process in (2.4), that include only a large enough number H of atoms, which we denote with $G_{\boldsymbol{u}_{i,t}}^H$. For details on the finite approximation of (2.4) see the work in Barcella et al. (2016). Then, the full conditional follows:

$$\Pr((\boldsymbol{\alpha}_{i,t}, \Sigma_{i,t}) = (\boldsymbol{\alpha}_h^*, \Sigma_h^*) \mid \ldots) \propto w_{h,\boldsymbol{u}_{i,t}} f(\boldsymbol{z}_{i,t} \mid \boldsymbol{\alpha}_h^*, \Sigma_h^*, \ldots), \text{ for } h = 1, \ldots, H.$$

v) *Sample $G_{\boldsymbol{u}_{i,t}}^H$ for all values of $\boldsymbol{u}_{i,t}$ given all other parameters.* This step consists of resampling the locations $\{(\boldsymbol{\alpha}_h^*, \Sigma_h^*), h = 1, \ldots, H\}$ and the weights $\{w_{h,\boldsymbol{u}_{i,t}}, h = 1, \ldots, H$ and all $\boldsymbol{u}_{i,t}\}$. For location parameters, full conditionals can be derived using the likelihood of the observations associated to each location and the prior distribution of $\boldsymbol{\alpha}_h$ and Σ_h (see Jara et al., 2007). Sampling the weights requires the joint distribution of the weights (marginally for each $\boldsymbol{u}_{i,t}$) which follow a generalised Dirichlet distribution (Connor and Mosimann, 1969). We augment the parameter space with clustering indicators $s_{i,t}$ taking value in $\{1, \ldots, H\}$. $s_{i,t}$ denotes which location of $G_{\boldsymbol{u}_{i,t}}^H$ the observation (i, t) is assigned to in step iv). It is to show that the $s_{i,t}$ follow a Multinomial distribution for $s_{i,t}$, which is conjugate to the generalised Dirichlet distribution, allowing easy determination of the full conditional of the weights.

vi) *Sample $(\boldsymbol{\mu}, \phi)$ given all other parameters.* This step requires a Metropolis scheme for sampling from the following full conditional

$$p(\boldsymbol{\mu}, \phi \mid \ldots) \propto \prod_{\boldsymbol{u}} p(\boldsymbol{w_u} \mid \boldsymbol{\mu}, \phi) p(\boldsymbol{\mu}, \phi),$$

where $\boldsymbol{w_u} = (w_{1,\boldsymbol{u}}, \ldots, w_{H,\boldsymbol{u}})$ and $p(\boldsymbol{w_u} \mid \boldsymbol{\mu}, \phi)$ is the generalised Dirichlet distribution.

References

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, 88(422):669–679.

Antoniak, C. E. (1974). Mixtures of dirichlet processes with applications to bayesian nonparametric problems. *The Annals of Statistics*, 1:1152–1174.

Ashford, J. and Sowden, R. (1970). Multi-variate probit analysis. *Biometrics*, 26(3):535–546.

Barcella, W., De Iorio, M., Favaro, S., and Rosner, G. L. (2016). Dependent generalised Dirichlet process priors for the analysis of acute lymphoblastic leukaemia. *Biostatistics*, In press.

Chen, M.-H. (2004). *Skewed link models for categorical response data*, pages 131–151. Chapman and Hall/CRC.

Chib, S. and Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, 85(2):347–361.

Connor, R. J. and Mosimann, J. E. (1969). Concepts of independence for proportions with a generalization of the Dirichlet distribution. *Journal of the American Statistical Association*, 64(325):194–206.

DeYoreo, M. and Kottas, A. (2017a). Bayesian nonparametric modeling for multivariate ordinal regression. *Journal of Computational and Graphical Statistics*, In press.

DeYoreo, M. and Kottas, A. (2017b). Modeling for dynamic ordinal regression relationships: an application to estimating maturity of rockfish in california. *Journal of the American Statistical Association*, In press.

DeYoreo, M., Kottas, A., et al. (2015). A fully nonparametric modeling approach to binary regression. *Bayesian Analysis*, 10(4):821–847.

Di Lucca, M. A., Guglielmi, A., Müller, P., and Quintana, F. A. (2013). A simple class of Bayesian nonparametric autoregression models. *Bayesian Analysis*, 8(1):63–88.

Dunson, D. B. and Park, J.-H. (2008). Kernel stick-breaking processes. *Biometrika*, 95(2):307–323.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230.

Giardina, F., Guglielmi, A., Quintana, F. A., and Ruggeri, F. (2011). Bayesian first order auto-regressive latent variable models for multiple binary sequences. *Statistical Modelling*, 11(6):471–488.

Hatjispyros, S. J., Nicoleris, T., and Walker, S. G. (2015). Dependent random density functions with common atoms and pairwise dependence. *arXiv preprint arXiv:1510.07153*.

Holmes, C. C. and Held, L. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis*, 1(1):145–168.

Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453):161–173.

Ishwaran, H. and Zarepour, M. (2000). Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika*, 87(2):371–390.

Jara, A., García-Zattera, M. J., and Lesaffre, E. (2007). A Dirichlet process mixture model for the analysis of correlated binary responses. *Computational Statistics & Data Analysis*, 51(11):5402–5415.

Kalli, M., Griffin, J. E., and Walker, S. G. (2011). Slice sampling mixture models. *Statistics and Computing*, 21(1):93–105.

Kottas, A., Müller, P., and Quintana, F. (2005). Nonparametric Bayesian modeling for multivariate ordinal data. *Journal of Computational and Graphical Statistics*, 14(3):610–625.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. density estimates. *The Annals of Statistics*, 12(1):351–357.

MacEachern, S. N. (1999). Dependent nonparametric processes. In *ASA proceedings of the section on Bayesian statistical science*, pages 50–55.

MacEachern, S. N. (2000). Dependent Dirichlet processes. Technical report, Department of Statistics, The Ohio State University.

O'Brien, S. M. and Dunson, D. B. (2004). Bayesian multivariate logistic regression. *Biometrics*, 60(3):739–746.

Ohlssen, D. I., Sharples, L. D., and Spiegelhalter, D. J. (2007). Flexible random-effects models using Bayesian semi-parametric models: applications to institutional comparisons. *Statistics in Medicine*, 26(9):2088–2112.

Papageorgiou, G., Richardson, S., and Best, N. (2015). Bayesian non-parametric models for spatially indexed data of mixed type. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(5):973–999.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika*, 86(3):677–690.

Ren, L., Du, L., Carin, L., and Dunson, D. (2011). Logistic stick-breaking process. *The Journal of Machine Learning Research*, 12:203–239.

Rodriguez, A. and Dunson, D. B. (2011). Nonparametric Bayesian models through probit stick-breaking processes. *Bayesian analysis*, 6(1):145–177.

Sethuraman, J. (1991). A constructive definition of dirichlet priors. Technical report, DTIC Document.

Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics—Simulation and Computation*, 36(1):45–54.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57(298):348–368.