Bayesian Analysis and Prediction of Patients' Demands for Visits in Home Care

Inad Nawajah[1], Raffaele Argiento[2],
Alessandra Guglielmi[1], Ettore Lanzarone[2]

Bayesian Young Statisticians Meeting (BAYSM), Milan June, 5-6, 2013
Paper no. 18

[1] Dipartimento di Matematica, Politecnico di Milano, Milan, Italy
alessandra.guglielmi@polimi.it
inad.nawajah@mail.polimi.it
[2] CNR–IMATI, Milan, Italy
raffaele@mi.imati.cnr.it
ettore@mi.imati.cnr.it

Abstract

Home Care (HC) providers are complex structures which include medical, paramedical and social services delivered to patients at their domicile. High randomness affects the service delivery, mainly in terms of unplanned changes in patients' conditions, which make the amount of required visits highly uncertain. In this paper, we propose a Bayesian model to represent the HC patient's demand evolution over time and to predict the demand in future periods. Results from the application to a relevant real case validate the approach, since low prediction errors are found.
Keywords: Home Care; Bayesian Modeling and Estimation; MCMC algorithm; Random effect.

1 Introduction

Home Care (HC) refers to any type of care provided to a patient at his/her own home. The main benefit of HC is the reduction of the hospitalization rate, which significantly increases the quality of life for the assisted patients and determines a relevant cost saving for the entire health care system (see [1]). Appropriate resource planning is required in HC for avoiding process inefficiencies and overloaded operators; in addition, many random events affect the service delivery and mine the feasibility of plans (see, for instance, [2, 3, 4]). Among them, the most frequent and critical event is a variation in patient's condition, which makes the demand for visits different from the planned one.

In the literature, several studies deal with stochastic models for representing patient conditions in the health care system and, among them, there are also Bayesian approaches. Moreover, Bayesian models have been used to predict patient traffic from their home to hospital, in order to facilitate reconfigurations of the emergency hospital services [5, 6].

With regard to HC, a patient stochastic model is proposed in [7]: patient evolution is described by means of a Markov chain, in which transition probabilities are derived with a frequentistic approach, and the demand for visit is obtained with a frequentistic cost function assigned to each state of the chain. To the best of our knowledge, Bayesian approaches have not been considered in the HC context so far. Therefore, the aim of this work is to propose a Bayesian model that represents and predicts the demand evolution of HC patients.

2 Bayesian Model

We consider m patients in charge of a HC provider over a period divided into discrete time slots. Each patient i enters the service at time slot $T_L(i)$ and exits at time $T_U(i)$. Data observed for each patient i at time slot $t \in [T_L(i), \dots, T_U(i)]$ are:

- $N_{i,t}$: number of visits required to nurses (counts data) by patient i at time slot t.
- $CP_{i,t}$: Care Profile of patient i at time slot t. CP a categorical covariate assuming N_{cp} integer values s (with $s = 1, \dots, N_{cp}$), evolving in time, assigned by the provider based on the specific requirements and the costs associated with the provided services. Usually, a CP is assigned to each patient at the beginning of the care pathway and monthly confirmed or changed. However, CP can be modified in advance in case of a sudden variation in patient conditions.

Moreover, patient i is characterized by sex_i (gender - categorical variable) and age_i (age at $t = T_L(i)$ – continuous positive variable).

We model each $N_{i,t}$ as a discrete Poisson distribution with expected value $\lambda_{i,t}$. The evolution of the latent variable $\lambda_{i,t}$ over t is determined according to a Markov chain. Let $\mathbf{N_i} = \left(N_{i,T_L(i)}, N_{i,T_L(i)+1}, N_{i,T_L(i)+2}, \dots, N_{i,T_U(i)}\right)$ for each i, and assume that $\mathbf{N_1}, \dots, \mathbf{N_m}$ are conditionally independent. We propose a generalized linear model as follows:

$$
\begin{aligned}
N_{i,t}|\lambda_{i,t} &\sim \text{Pois}\left(\lambda_{i,t}\right), \qquad T_L(i) \le t \le T_U(i) \\
\log\left(\lambda_{i,t}\right) &\sim N\left(\alpha[CP_{i,t}] \log\left(\lambda_{i,t-1}\right) + \beta\left[CP_{i,t}\right], \sigma^2\right), \quad T_L(i) < t \le T_U(i) \\
\log\left(\lambda_{i,T_L(i)}\right) &\sim N\left(\gamma_1 age_i + \gamma_2 sex_i + \gamma_3\left[CP_{i,T_L(i)}\right], \sigma_0^2\right).
\end{aligned}
$$

In this formulation, the latent variable $\lambda_{i,t}$ represents the health status of patient i in time slot t, which is responsible for his/her demand for visits (the bigger the parameter $\lambda_{i,t}$ is, the worse the patient's conditions are), while $CP_{i,t}$ is a fixed covariate here (in this paper), and parameters $\alpha_s = \alpha[CP_{i,t} = s]$, $\beta_s = \beta[CP_{i,t} = s]$ and $\gamma_{3,s} = \gamma_3[CP_{i,t} = s]$ describe the random-effects as a function of the assumed $CP_{i,t}$. The model we proposed here is a generalization of the model in [8]. Parameters $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma_1, \gamma_2, \boldsymbol{\gamma}_3, \sigma^2, \sigma_0^2)$ are a priori (conditionally) independent and their marginal prior densities are listed below:

$$\begin{aligned}
\alpha_s &\overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\alpha^2), \quad s = 1, \ldots, N_{cp} \\
\beta_s &\overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\beta^2), \quad s = 1, \ldots, N_{cp}, \\
\gamma_{3s} &\overset{\text{iid}}{\sim} \sim\mathcal{N}(0, \sigma_{\gamma_3}^2), \quad s = 1, \ldots, N_{cp}, \\
(\gamma_1, \gamma_2) &\sim \mathcal{N}_2(0, 1000), \sigma^2 \sim U(0, 5),
\end{aligned}$$

and σ_0^2 has a fixed value equal to three. Moreover, $\sigma_\alpha, \sigma_\beta$ and σ_{γ_3} independent, where

$$\sigma_\alpha \sim U(0, 5), \sigma_\beta \sim U(0, 2) \text{ and } \sigma_{\gamma_3} \sim U(0, 15).$$

The Bayesian formulation seems very appropriate, since, under this approach, prediction of $N_{i,t+1}$ at the next time slot is naturally accommodated by means of predictive distribution. This is very important for HC decision makers, who are interested in assigning nurses to patient over a future planning horizon to improve service efficiency. The predictive distribution is:

$$\begin{aligned}
&\mathcal{L}\left(N_{i,t+1} = k | covariate, \mathbf{N_1}, \ldots, \mathbf{N_m}\right) = \\
&= \int \mathcal{L}\left(N_{i,t+1} = k | \lambda_{i,t+1}\right) \mathcal{L}\left(\mathrm{d}\lambda_{i,t+1} | \lambda_{i,t}\right) \pi\left(\mathrm{d}\lambda_{i,t} | \mathbf{N_1}, \ldots, \mathbf{N_m}\right) \qquad (1)
\end{aligned}$$

where $N_{i,t+1}$ is the number of nurse visits at time slot $t+1$.

From a predictive point of view, the accuracy of the prediction will be evaluated in terms of the Mean Absolute Error (MAE), that is:

$$MAE_{t+1} = \frac{\sum_{i=1}^{m} |n_{i,t+1} - \hat{N}_{i,t+1}|}{m_t},$$

where m_t is the number of patients in charge at week t, $n_{i,t+1}$ is the observed number of nurse visits at time slot $t+1$, and $\hat{N}_{i,t+1}$ is the defined as the mode of the predictive distribution in (1). In this way, $\hat{N}_{i,t+1}$ is assumed as the Bayesian prediction to be compared with the real observation.

3 Application to real data and results

We apply the model to one of the largest Italian home care providers; a description of the data set and its frequentist analysis are reported in [7]. Here, the week is considered as the time slot; 252 weeks (from 2004 to 2008) are included in the study. The provider consists of three divisions and we refer to patients of the largest one. Moreover, we considered only patients who entered and left the service once within our time window. In this way, our dataset consists of 2401 patients.

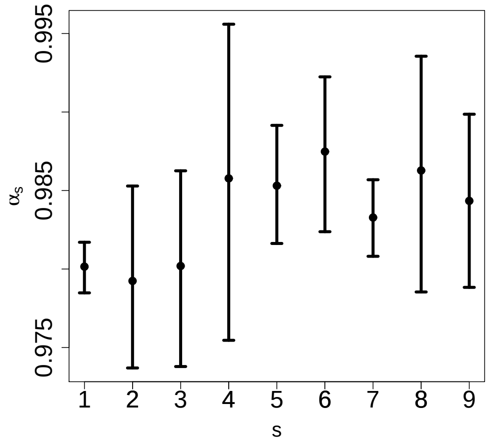
Patients are grouped in two categories (palliative and non palliative patients), and each category includes a certain number of CPs. Fifteen CPs are present in the provider [7]. However, after joining together very similar CPs, we reduce the number to 9 (see Table 1).

Table 1: Classification of CPs.

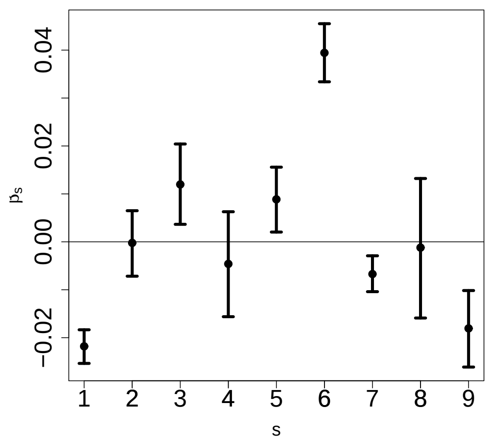
Type of care	CPs of the provider	Our group
Extemporary Care with a very low frequency of visits	1 15	1 9
Integrated Home Care characterized by a medium–high care intensity (CP are listed in increasing order of expected number of weekly visits)	2, 12 3, 13 4, 14 5 9 10	2 3 4 5 7 8
Palliative Care for terminal patients generally affected by oncological diseases (CPs are listed in increasing order of expected number of weekly visits)	6,7,8	6

In order to compute the Bayesian estimates, the model was implemented in Jags ([9]), with chains consisting of 250000 iterations with a burn-in of 5000 and a thinning of 50 iterations. The chains passed most of the standard convergence tests.

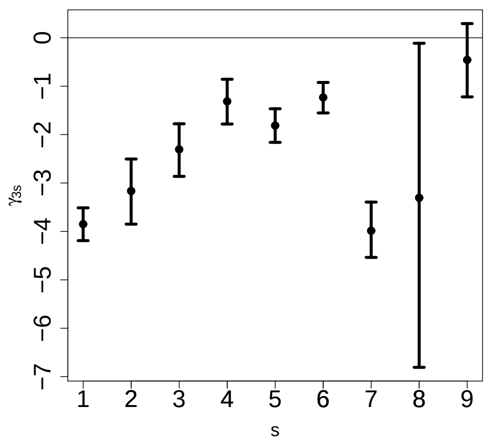
Figures 1 reports the posterior credibility intervals of the main parameters of our model, corresponding to CP equals to $1, \ldots, 9$, respectively.



(a)



(b)



(c)

Figure 1: 95% credible intervals of α_s, β_s and $\gamma_{3s}, s = 1, 2, ..., 9$.

In particular, the estimated values of $\beta[CP]$ parameters are clearly negative when CP is 1 or 9, clearly positive for CP= 6, and closer to 0 for the other CP values, leading to a subdivision coherent with the one described in Table 1.

From a predictive point of view, we computed the predicted nurse visits of the active patients at week $t + 1$ (with, $t = 99,\ 149,\ 175,\ 234$). The estimated MAE are given in Table 2.

Table 2: The MAE at week $t+1$.

$t+1$	100	150	176	235
MAE	0.52	0.63	0.65	0.55

The largest MAE is 0.65 at week 176, showing a very good fit of the model to the analyzed data. To check if there is a systematic error in on our prediction, we examined the error plot, where

$$Error_{t+1} = n_{it+1} - \hat{N}_{it+1}$$

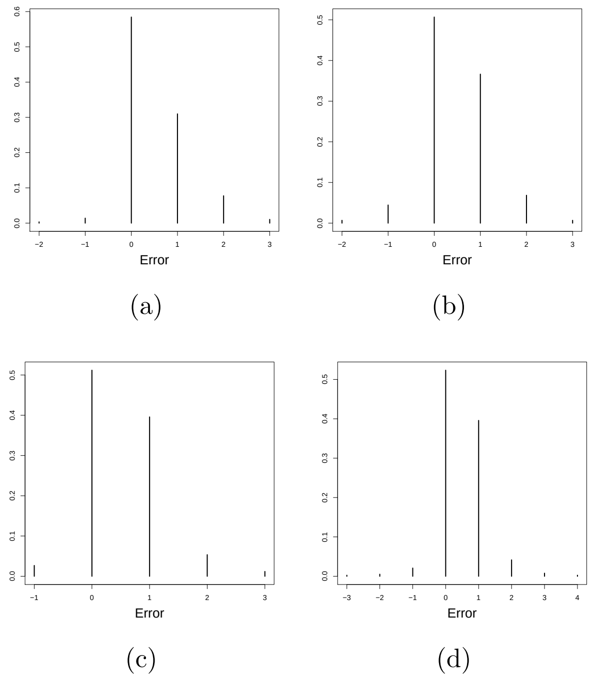


Figure 2: Error plots for prediction at week 100 (a), week 150 (b), week 176 (c) and week 235 (d).

Figure 2 shows that our predictive estimates are mostly underestimating the number of visits effectively given to the patients, although the absolute value of the error is relatively small and most frequently equal to 1.

4 Conclusion

In this work, we first explored the application of a Bayesian models to the HC context, in order to predict the demand for visits from patients in charge. The approach fits well the HC context, and the results from the application to a relevant real case validate the approach, since low prediction errors are found. Hence, the applicability of the proposed model in the practice is guaranteed. Our future work will deal with joint Bayesian estimation of the number of visits and the CPs in future periods.

References

[1] V. Comondore, P. Devereaux, Q. Zhou, et al. **Quality of care in for-profit and not-for-profit nursing homes: systematic review and meta-analysis**. *British Medical Journal*; 2009; 339; b2732.

[2] E. Lanzarone, A. Matta, E. Sahin. **Operations management applied to home care services: the problem of assigning human resources to patients**. *IEEE Transations on System Man Cybernetics A* 2012; 42(6): 1346-63.

[3] E. Lanzarone, A. Matta. **A cost assignment policy for home care patients**. *Flexible Service and Manufacturing Journal*; 2012; 24(4): 465-495.

[4] A. Matta, S. Chahed, E. Sahin, Y. Dallery. **Modeling home care organizations from an operations management prospective**. *Flexible Service and Manufacturing Journal*; doi:10.1007/S10696-012-9157-0.

[5] P. Congdon. **The development of Gravity models for hospital patient flows under system change:A Bayesian modelling approach**. *Health Care Managment Science*; 2001; 4; 289-304.

[6] A. Marshall, C. Vasilakis, E. El-Darzi. **Length of stay–based patient flow models: recent developments and future directions**. *Health Care Management Science*; 2005; 8; 213-220.

[7] E. Lanzarone, A. Matta, G. Saccabarozzi. **A patient stochastic model to support human resource planning in home care**. *Production Planning and Control*; 2010; 21; 3-25.

[8] F. Giardina, A. Guglielmi, F. Quintana, F. Ruggeri. **Bayesian first order auto-regression latent variable models for multiple binary sequences**. *Statistical Modeling*; 2001; 11(6); 471-488.

[9] Plummer, M. (2003). **JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling**. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, 20–22.