Analysis of hospitalizations of patients affected by chronic heart disease

Alice Parodi[1], Francesca Ieva[1], Alessandra Guglielmi[1], Raffaele Argiento[2]

Bayesian Young Statisticians Meeting (BAYSM), Milan June, 5-6, 2013
Paper no. 29

[1] Mathematical Department, Politecnico di Milano, Milano, Italy
`alice.parodi@mail.polimi.it`, `francesca.ieva@polimi.it`,
`alessandra.guglielmi@polimi.it`
[2] CNR-IMATI, Milano, Italy
`raffaele@mi.imati.cnr.it`

1 Introduction

In this paper we present a preliminary analysis of the evolution of a chronicle heart disease [section 2] considering as variables of interest the sequence of hospitalizations of patients affected by this illness and times of these events [1].
In particular we introduce a model [section 3] with the scope of evaluating factors that are relevant in the evolution of the sequence of hospitalizations and in the time of registration of these events.

2 Presentation of dataset

The dataset contains informations about hospitalizations of patients affected by chronic heart disease. In particular for each patient $j = 1, \dots G$ we consider the sequence of hospitalizations and the period occurred between subsequent events. The analysis is carried out on $G = 26.618$ patients examined from 1/1/2006 to 31/12/2010. For each of them we know the number n_j of events A_i^j (hospitalization or death) occurred and the sequence of times observed between subsequent events (say T_i^j the time between A_{i-1}^j and A_i^j). We are also informed about covariates that possibly influence the occurrence of events: i.e. clinical covariates like status (0 healthy, 1 not), age and sex (1 female, 0 male).

3 Presentation of the model

To describe the model we assume patients to be independent. Further specifications necessary to fit the model are the following:

h1: Each patient has at least one hospitalization $A_1^j = 1$

h2: Following events can be hospitalization or death $A_i^j = i$ or $A_i^j = M$ with $i \in \{2, \ldots, n_j\}$

h3: The event A_i^j depends on all the previous times, but only on the last event

h4: The time T_i^j depends on all the previous times and on the last event only, i.e. T_i^j is independent from A_k^j $\forall k < i-1$

h5: If a patient is not dead before the end of the analysis we can define time $T_{n_j+1}^j$ as the time of the first event occurred after the end of the study.

Observation: For the preliminary analysis we considerer only patients experiencing at most $N = 10$ hospitalizations, in order to guarantee datas include enough information for all events so that all parameters can be inferred.

Then we can define the likelihood of the model:

$$\mathcal{L}(\underline{T}, \underline{A}|\underline{\rho}) = \prod_{j=1}^{G} \mathcal{L}((T_1^j, A_1^j), (T_2^j, A_2^j), \ldots, (T_{n_j}^j, A_{n_j}^j), T_{n_j+1}^j|\underline{\rho})$$

and, at for each patient $j \in \{1, \ldots, G\}$

$$\begin{aligned}
\mathcal{L}(\underline{T^j}, \underline{A^j}|\underline{\rho}) \quad = \quad & \mathcal{L}(T_1^j|\underline{\rho}) \quad \cdot \\
\cdot \quad & \mathcal{L}(A_2^j|T_1^j, T_2^j, A_1^j = 1, \underline{\rho}) \;\; \mathcal{L}(T_2^j|T_1^j, A_1^j = 1, \underline{\rho}) \;\; \cdot \\
\cdot \quad & \mathcal{L}(A_3^j|T_1^j, T_2^j, T_3^j, A_2^j, \underline{\rho}) \;\; \mathcal{L}(T_3^j|T_1^j, T_2^j, A_2^j, \underline{\rho}) \;\; \cdot \\
\cdot \quad & \mathcal{L}(A_4^j|T_1^j, \ldots, T_4^j, A_3^j, \underline{\rho}) \;\; \mathcal{L}(T_4^j|T_1^j, T_2^j, T_3^j, A_3^j, \underline{\rho}) \;\; \cdot \\
\cdot \quad & \ldots\ldots \; \cdot \\
\cdot \quad & \mathcal{L}(A_{n_j}^j|T_1^j, T_2^j, \ldots, T_{n_j}^j, A_{n_j-1}^j, \underline{\rho}) \;\; \cdot \\
& \qquad\qquad \cdot \;\; \mathcal{L}(T_{n_j}^j|T_1^j, \ldots, T_{n_j-1}^j, A_{n_j-1}^j, \underline{\rho}) \;\; \cdot \\
\cdot \quad & \mathcal{L}(T_{n_j+1}^j|T_1^j, \ldots, T_{n_j}^j, A_{n_j}^j, \underline{\rho})
\end{aligned}$$

Now, defining specifically each term, we have:

- $(T_1^j|\underline{\rho}) \sim \text{Weibull}(\lambda_1, \mu_1^j)$ being $\mu_1^j = e^{\underline{z}_j^T \underline{\gamma}}$ and $\underline{z}_j = (1, \text{age}_j, \text{sex}_j)$; $\underline{\gamma} = (\gamma_0, \gamma_1, \gamma_2)$
- $\mathcal{L}(A_2^j|T_1^j = t_1, T_2^j = t_2, A_1^j = 1, \underline{\rho}) = \begin{cases} p_2 & \text{if } A_2^j = 2 \\ 1 - p_2 & \text{if } A_2^j = M \end{cases}$
  We assume the probability $p_2 = p_2(t_1, t_2) \sim \text{Beta}(\alpha_2(t_1, t_2) q_2(t_1, t_2), \alpha_2(t_1, t_2)(1 - q_2(t_1, t_2)))$ and we define $\alpha_2(t_1, t_2) = \frac{a}{t_1 \, t_2}$ and $q_2(t_1, t_2) = q_2$

- $(T_2^j | T_1^j = t_1, A_1^j = 1, \underline{\rho}) \sim \text{Weibull}(\lambda_2, \mu_2^j)$
  where $\mu_2^j = e^{\underline{\beta}^T \underline{x}_2^j}$ and $\underline{x}_2^j = (1, \text{age}_j, \text{sex}_j, \text{cov}_2^j, t_1)$, $\underline{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$

  Note that the definition of the parameter β_4 is ancillary to introduce into the law of T_2^j the explicit dependence on the past.

- $\mathcal{L}(A_3^j | T_1^j = t_1, T_2^j = t_2, T_3^j = t_3, A_2^j = 2, \underline{\rho}) = \begin{cases} p_3 & \text{if } A_3^j = 3 \\ 1 - p_3 & \text{if } A_3^j = M \end{cases}$
  where $p_3 = p_3(t_1, t_2, t_3) \sim \text{Beta}(\alpha_3(t_1, t_2, t_3) q_3(t_1, t_2, t_3), \alpha_3(t_1, t_2, t_3)(1 - q_3(t_1, t_2, t_3)))$
  and then $\alpha_3(t_1, t_2, t_3) = \frac{a}{(t_1+t_2)\, t_3}$ e $q_3(t_1, t_2, t_3) = q_3$

  According to the meaning of each state A_i^j we consider only $A_2^j = 2$; if $A_2^j = M$ the event A_3^j will not occur. Infact for each patient j we know the total number of events n_j.

- $(T_3^j | T_1^j = t_1, T_2^j = t_2, A_2^j = 2, \underline{\rho}) \sim \text{Weibull}(\lambda_3, \mu_3^j)$
  where $\mu_3^j = e^{\underline{\beta}^T \underline{x}_3^j}$ and $\underline{x}_3^j = (1, \text{age}_j, \text{sex}_j, \text{cov}_3^j, t_1 + t_2)$ and$\underline{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$

- $(T_{n_j+1}^j | T_1^j = t_1, \ldots, T_{n_j}^j = t_{n_j}, A_{n_j}^j = n_j, \underline{\rho}) \sim \frac{\text{Weibull}(\lambda_{n_j}, \mu_{n_j}^j)}{S(\overline{T} - \sum_{i=1}^{n_j} t_i)} \, \mathbb{I}_{(\overline{T} - \sum_{i=1}^{n_j} t_i; +\infty)} \left\{ T_{n_j+1}^j \right\}$
  where $\mu_{n_j}^j = e^{\underline{\beta}^T \underline{x}_{n_j}^j}$ and $\underline{x}_{n_j}^j = (1, \text{age}_j, \text{sex}_j, \text{cov}_{n_j}^j, \sum_{k=1}^{n_j} t_k)$, $\underline{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$

To complete the definition of the model we introduce the prior distribution of parameters, in particular we have:

$$\pi(\underline{\rho}) = \pi(\underline{\beta})\, \pi(\underline{\gamma})\, \pi(\underline{\lambda})\, \pi(a)\, \pi(\underline{q})$$

where: $\pi(\underline{\beta}) = \prod_{k=0}^{4} \pi(\beta_k)$ $\quad \beta_k \sim N(0, 1000) \quad \forall k \in \{0, 1, \ldots, 4\}$
$\pi(\underline{\gamma}) = \prod_{k=0}^{2} \pi(\gamma_k) \quad \gamma_k \sim N(0, 1000) \quad \forall k \in \{0, 1, 2\}$
$\pi(\underline{\lambda}) = \prod_{k=1}^{N-2} \pi(\lambda_k) \quad \lambda_k \sim \text{Gamma}(\eta, \nu)$ and $\eta = 10$, $\nu = 5$; $\forall k \in \{1, 2, \ldots, N-1\}$
and $\lambda_N = \lambda_{N-1} = \lambda_{N-2}$
$\pi(a) = \text{Gamma}(2, 4)$
$\pi(\underline{q}) = \prod_{k=2}^{N-2} \pi(q_k) \quad q_k \sim U(0, 1)$, $\forall k \in \{2, 3, \ldots, N-1\}$ and $q_N = q_{N-1} = q_{N-2}$

4 Analysis of results

In this section we present the analysis of results obtained with a Gibbs Sampling algorithm runned over 100.000 iterations with thinning of 10 iterations discarding 1000 iterations for burn-in (model implemented with JAGS [2]; [3]).

Concernig the distribution of time, chains connected to parameters $\underline{\beta}$, $\underline{\gamma}$ and $\underline{\lambda}$ converge well and are consistent with different initialization of parameters. In figure 1 we present the chain connected to parameter β_1. All the parameters introduced before have similar posterior results.

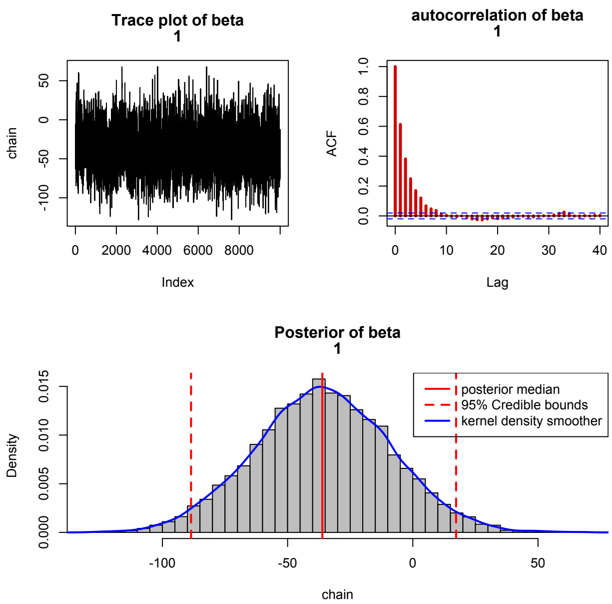

Figure 1: Analysis of β_1

Therefore we can consider numerical values of the parameters and we observe that the *shape* parameters of the Weibull distribution have values around 1. Regression parameters can also be explained, in particular coefficients connected to age have values around 0 and (after specific test of validation of the model) we can conclude that the explicit influence of age is negligible; about sex and clinical situation we observe that women and unhealthy patients have shorter time for new hospitalization. We conclude also that the longer is the time from the first hospitalization, the longer the time to a new event would be.

	mean	SD	2.5%	25%	50%	75%	97.5%
β_0	-35.10	27.67	-89.91	-53.98	-34. 88	-16.12	19.04
β_1	-1.15	0.40	-1.91	-1.42	-1.15	-0.88	-0.36
β_2	-21.82	13.06	-47.34	-30.66	-21.85	-13.02	3.50
β_3	-16.14	4.03	-27.06	-18.83	-16.17	-13.39	-8.27
β_4	67.55	7.18	53.70	62.73	67.53	72.32	81.87

	mean	SD	2.5%	25%	50%	75%	97.5%
γ_0	-15.62	27.97	-70.00	-34.88	-15.45	3.06	39.89
γ_1	-1.08	0.42	-1.89	-1.36	-1.08	-0.80	-0.26
γ_2	-3.20	17.87	-38.06	-15.40	-3.29	9.02	31.85

Considering, instead, the law of the sequence of events we observe that neither the chain of a nor the chains of q_i converge well, so we can not infer anything about the sequence of events; we should introduce a new definition of the law of $A_i^j|\underline{\rho}$, for example without the explicit dependence on time.

5 Conclusion

As introduced before we can observe that the model presented analyzes well the evolution of time and defines well the relevant parameters. Considering, instead, the sequence of events we should change the model, for example we can introduce the explicit prior knowledge on the number of event (p_i is decreasing in i) and ignore the connection of p to time t.

These could be ideas for future developments of the model that should be detailed in further analysis.

References

[1] P. K. Andersen, N. Keiding, **Multi-state models for event history analysis**. *Statistical Methods in Medical Research*; 2002; 11; 91.

[2] D. Lunn, C. Jackson, N. Best, A.Thomas, D. Spiegelhalter, **The BUGS Book - A Practical Introduction to Bayesian Analysis** ; *CRC Press*; 2012.

[3] M. Plummer, **JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling**. *Title of the Journal*; 2012; 20(12); pp. 1-10.