# A semiparametric Bayesian multivariate model for survival probabilities after acute myocardial infarction

E. Prandoni[1], A. Guglielmi[1], F. Ieva[1], A.M. Paganoni[1],

*Bayesian Young Statisticians Meeting (BAYSM), Milan June, 5-6, 2013*
*Paper no. 35*

[1] Politecnico di Milano, Dipartimento di Matematica, Milano, Italy
alessandra.guglielmi@polimi.it
francesca.ieva@mail.polimi.it
anna.paganoni@polimi.it
elena.prandoni@mail.polimi.it

**Abstract**

In this work, a Bayesian semiparametric multivariate model is fitted to study data related to in-hospital and 60-days survival probabilities of patients admitted to a hospital with ST-elevation myocardial infarction diagnosis. We consider a hierarchical generalized linear model to predict survival probabilities and a process indicator (time of intervention). Poisson-Dirichlet process priors, generalizing the well known Dirichlet process, are considered for modeling the random-effects distribution of the grouping factor which is the hospital of admission.

**Keywords**: Bayesian nonparametrics; Generalized linear mixed models; Prediction.

## 1 Introduction

The disease we are interested in is ST-Elevation Myocardial Infarction (STEMI): it is caused by an occlusion of a coronary artery which causes an ischemia that, if untreated, can damage heart cells and make them die (infarction). It is very important that a reperfusion therapy could be done as quickly as possible, because its benefits decrease with delay in treatment; in our case, patients are treated with percutaneous transluminal coronary angioplasty. We consider data collected in the STEMI Archive [1], a multicenter observational prospective clinical study planned within the Strategic Program of Regione Lombardia. Data is recorded in a registry collecting clinical and process indicators, outcomes and personal information on patients admitted to all hospitals of Regione Lombardia with STEMI diagnosis.

The regression model we introduce is multivariate, where the response has three components: door to balloon time (DB), i.e. the time between the admission to the hospital and angioplasty, on the logarithmic scale, in-hospital survival and survival after 60 days from admission. The first term is an important indicator of the efficiency of health providers and plays a key role in the success of the therapy; the second is the basic indicator of success or failure of treatment, while the third is a very important outcome, since doctors believe that it is in a 60 days period the effectiveness of the treatment in terms of survival and quality of life can be truly evaluated. We include the hospital random-effect parameters in the model, and assume they are a sample from a Poisson-Dirichlet process a priori, in order to eventually cluster the hospitals.

The main statistical aim of this work is prediction of both survival probabilities of new patients.

## 2 The Bayesian model in a nutshell

For each patient ($i = 1, ..., 697$) let $\mathbf{Y}_i := (Y_{i1}, Y_{i2}, Y_{i3})$ be the response, where $Y_1$ is the logarithm of DB, $Y_2$ is the in-hospital survival and $Y_3$ is the long-term survival. We assume that observations, given parameters and covariates, are independent and the law of the response can be factorized in three parts:

$$\mathcal{L}(\mathbf{Y}_i|par, cov) = \mathcal{L}(Y_{i1}|par_1, cov_1)\mathcal{L}(Y_{i2}|Y_{i1}, par_2, cov_2)\mathcal{L}(Y_{i3}|Y_{i2}, par_3, cov_3).$$

The likelihood can be expressed as

$$Y_{i1}|\mu_i, \sigma \overset{ind}{\sim} \mathcal{N}(\mu_i, \sigma^2), \quad \mu_i = \sum_{l=1}^{4} \beta_l u_{il} + \beta_5 x_{i5} + \beta_6 x_{i6} \tag{1}$$

$$Y_{i2}|p_i, Y_{i1} \overset{ind}{\sim} Be(p_i), logit(p_i) = \alpha_1 z_{i1} + \alpha_2 z_{i2} + \alpha_3 Y_{i1} + \sum_{l=4}^{7} \alpha_l v_{il} + b_{\phi_{k[i]}k[i]} \tag{2}$$

$$Y_{i3}|r_i, Y_{i2} \overset{ind}{\sim} \begin{cases} Be(r_i) & \text{se } Y_{i2} = 1 \\ \delta_0 & \text{se } Y_{i2} = 0 \end{cases}, \quad logit(r_i) = \sum_{j=1}^{4} \gamma_j s_{ij} + t_{\phi_{k[i]}k[i]}. \tag{3}$$

Here $k[i]$ denotes the hospital where the patient $i$ is admitted to, while covariates includes: the type of rescue unit sent to the patient ($u_{il}$, $l = 1, \ldots, 4$), the time of the first ECG ($x_{i5}$), the age ($z_{i1}$), the Killip class ($v_{il}$, $l = 1, \ldots, 4$, which quantify in four categories the severity of infarction), other covariates measuring the health status of the patient, or if the treatment was successful or not. The indexes $\phi_{k[i]}$ of the random-effects parameters in (2) and (3) assumes values 1 or 0, if the hospital is in Milano or not.

As far as the fixed-effects parameters are concerned, we considered a parametric prior; for the random-effects parameters we assume Poisson-Dirichlet process priors with parameters $(f, g)$ [2]. This choice helps us to avoid dependency on parametric assumptions, to increase flexibility in the prior and more robust inferences. We obtained posterior estimates of all the parameters through a Gibbs sampler algorithm implemented in `JAGS` [3] and with the support of `R` [4]; for this purpose we used a truncated stick-breaking representation of the nonparametric prior, which is a generalization of the well known Dirichlet process [5].

## References


[1] F. Ieva, "Designing and mining a multicenter observational clinical registry concerning patients with Acute Coronary Syndromes" in New Diagnostic, Therapeutic and Organizational Strategies for Acute Coronary Syndromes Patients (eds: N.Grieco, M.Marzegalli, A.M.Paganoni), Springer 2013.

[2] J. Pitman, M. Yor. **The two-parameter PoissonDirichlet distribution derived from a stable subordinator**. *The Annals of Probability*; 1997; 2; pp. 855-900.

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

[4] R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna; 2009.

[5] J. Sethuraman. **A constructive definition of Dirichlet process prior**. *Statistica Sinica*; 1994; 2; pp. 639-650.