Joint modeling of Longitudinal outcomes and recurrent events using a Bayesian non-parametric Dirichlet Process prior approach

Azeez Adeboye*,
Supervisors: Prof. Y. Qin, Dr. J. Ndege and Dr. R. Mutambayi

Department of Statistics, University of Fort Hare, South Africa

**60th SASA Conference, University of South Africa,
November, 2018**

Contents

Introduction

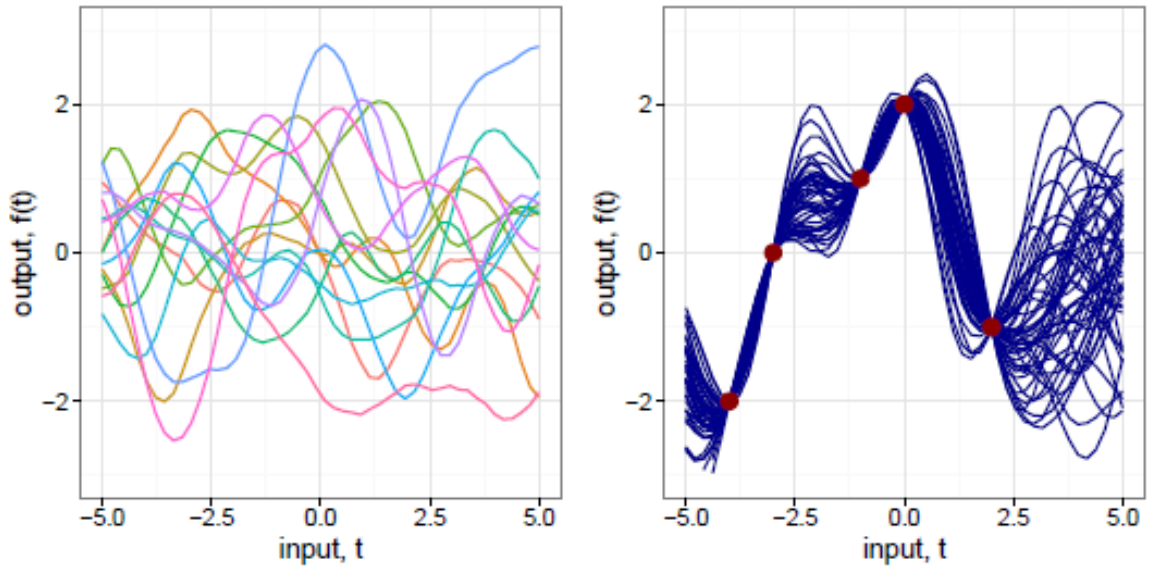
- In the medical research different kinds of patient information are gathered over time together with clinical outcome data such as overall survival.
- Joint models enable the analysis of correlated data of different types such as individual repeated data, clustered data together with overall survival.
- The repeated data may be recurrent events (e.g., relapses of a tumor, re-hospitalizations) or a longitudinal outcome called biomarker (e.g., tumor size, prostate-specific antigen or CD4 cell counts).
- A joint model for a longitudinal biomarker and a terminal event received the most of the attention in the literature. It estimates simultaneously the longitudinal and survival processes using the relationship via a latent structure of random-effects.
- A joint models for recurrent events and a terminal event (joint frailty models). The processes are linked via a random effect that represents the frailty of a subject (patient) to experience an event.

Model specification

Gaussian process (GP)

- Gaussian process (GP) can be thought as a Bayesian non-parametric technique that is widely used to define a prior distribution over functions
- It has been used in statistics literature for a long time (O'Hagan and Kingman (1978) ,Wahba (1990), Rasmussen (2006), Neal (2012)).
- It is useful in non-parametric Bayesian regression models to relax any explicit functional assumption using a prior distribution on functions (infinite-dimensional).
- It is considered as a regression with no explicit functional assumption and hence, with a great flexibility.
- Let $F = (f(t_1), (f(t_2),\ldots, (f(t_N)$ be an N-dimensional random vector of function values evaluated at *N* input points $t_i \in \Upsilon$, where $i \in \{1, \ldots, N\}$.
- A random function *f* is distributed according to a Gaussian process if for any finite subset $\{t_1, t_2, \ldots, t_N\} \subset \Upsilon$, *F* is distributed according to a multivariate Gaussian distribution.
- Gaussian processes are fully specified by a mean function, $\mu(t)$, and a covariance function, $C(t, t')$, where t and t' are two values from the input space as: $f(t) \sim GP\mu(t)C(t, t')$, where realizations of GP are random functions *f(t)*.

- Gaussian process prior example: we randomly sampled 15 functions from a Gaussian process with a squared exponential covariance function, where $\kappa^2 = 1, \rho^2 = 0.5$, and with an input space $\Upsilon = (-5, 5)$. Using observed data, out of all plausible functional forms under the specified GP prior, only those functions that are consistent with the observed data are selected. In case of no measurement error, consistent functions are functions that pass through all observed points.



- The left figure includes 15 randomly sampled functions from a Gaussian process with $\kappa^2 = 1, \rho^2 = 0.5$.
- The right figure includes samples from the posterior of the Gaussian process after observing data with a noise-free measurements.
- Posterior samples functions perfectly pass through all the data points.

Dirichlet Processes (DP)

- If parametric distribution has too many parameters, compared to the amount of data observed, a model may suffer from over-fitting or under-fitting if there are not enough parameters to model the data.
- A proper model selection technique is needed, which is often not an easy task in parametric models. A Bayesian non-parametric techniques with an unbounded number of parameters, where the posterior samples of the parameters used to model data, can avoid both under-fitting and over-fitting.
- The Dirichlet process (DP) is a Bayesian non-parametric technique that is often used to create flexible models that allow for a broad class of distributions.
- It is often used in density estimation without any explicit parametric distributional assumptions.
- The Dirichlet process was first introduced by Ferguson (1973) as a generalization of the Dirichlet distribution to infinite-dimensional space, where he proposed a Dirichlet process construction using a normalized Gamma process.
- Consider a random variable *X* and *n* iid sampled values of the form $x_i, x, \dots, x_n$ from the sample space Ω. Random variable *X* is distributed according to an unknown distribution *G (X|G ~ G)*, where *G* has a Dirichlet process prior ($G \sim DP(\alpha, G_0)$). The posterior distribution of *G*, given the observed data, can be written as:

$$G|x_i, x, \dots, x_n \sim DP(\alpha + n, \frac{\alpha}{\alpha + n} G_0 + \frac{1}{\alpha + n} \sum_{i=1}^{n} \delta_{x_i}$$

Dirichlet process mixture prior used in Longitudinal data

- We propose a hierarchical Bayesian model capable of detecting latent subgroup effects that are in the form of latent random intercepts.
- The models is capable of estimating conditional parameters.
- The proposed model is robust to distributional misspecification of the random intercepts and as opposed to an explicit distributional assumption.
- Further, DPM prior will allow subjects to cluster based on the distributional similarities of their latent random intercepts, hence, provides higher precision in estimating those latent subject effects

$$X_i \mid \beta_{0i}^{(\mathrm{L})}, \kappa_i, \rho, \sigma^2 \sim N(\beta_{0i}^{(\mathrm{L})}, \kappa_i K_i + \sigma^2 I_{l_i \times l_i}),$$

where $\beta_{0i}^{(L)}$ = subject-specific random intercept for subject *i;* κ_i = subject-specific measure of volatility in the longitudinal biomarker for individual; ρ^2= fixed correlation length; σ^2=measurement error shared across all subjects

as a Mean-DPM model

$$\beta_{0i} \sim N(\mu_i, \sigma_{\beta_{0i}}),$$
$$\mu_i \mid G \sim G,$$
$$G \sim \mathrm{DP}(\alpha, \mathrm{G}_0 = N(0, \sigma_0))$$

as Sigma-DPM model

$$\beta_{0i} \sim N(\mu_i, \sigma_{\beta_{0i}}^{(i)}),$$
$$\sigma_{\beta_{0i}}^{(i)} \mid G \sim G,$$
$$G \sim \mathrm{DP}(\alpha, \mathrm{G}_0 = \log\text{–}Normal(\mu \mathrm{G}_0, \sigma \mathrm{G}_0))$$

Proportional Hazard model

- Propose a hierarchical Bayesian proportional hazards model capable of detecting the differential subject-specific baseline hazard risk across subjects.

$$\lambda(T_i \mid \mathrm{Z}_i^{(s)}) = \lambda_0(T_i)\exp(\psi\, \mathrm{Z}_i^{(s)}(t)),$$

We consider a Weibull distribution for the survival component to allow for log-linear changes in the baseline hazard function over time

$$T_i \mid \tau, \psi, \mathrm{Z}_i^{(s)} \sim Weibull(\tau, \lambda_i = \beta_{0i}^{(s)} + \psi\, \mathrm{Z}_i^{(s)}),$$

- A two-stage approach to associating biomarker volatility with the survival outcome has been proposed by Holsclaw et al. (2014). As a comparison, two-step Cox model where the longitudinal curve of TB biomarker and its derivative curve are estimated using hyper-parameters set as the posterior median of a Bayesian Gaussian Process model.

where T_i = survival time; τ = shape parameter of the Weibull distribution; $\Psi^{(s)}$ = vector of coefficients relating baseline survival covariates to the risk of the occurrence of the event of interest; λ_i = log of the scale parameter in the Weibull distribution; $\beta_{0i}^{(s)}$ = subject-specific baseline hazard for subject i, Z_i = vector of survival coefficients.

Flexible joint model for this study

Model I: ***a Survival model with Longitudinal Biomarker at event time as a covariate.***

- Association between a longitudinal biomarker of interest and survival outcome by adjusting for the biomarker values in the survival component.

$$T_i \mid \tau, \psi^{(s)}, \psi_{X_i} \sim Weibull(\tau, \lambda_i = \beta_{0i}^{(s)} + \psi^{(s)} \mathrm{Z}_i^{(s)} + \psi_{X_i} X_i(t)),$$

Model II: ***A Survival model with covariates of Biomarker Value and the Derivative of its Trajectory at event Time***

- Model I is extended by including a measure of the average slope of the biomarker over time through derivative of the posterior mean trajectory of the biomarker with respect to time.

$$T_i \mid \tau, \psi^{(s)}, \psi_{X_i}, \psi_{X_i^*} \sim Weibull(\tau, \lambda_i = \beta_{0i}^{(s)} + \psi^{(s)} \mathrm{Z}_i^{(s)} + \psi_{X_i} X_i(t) + \psi_{X_i^*} X^*_{ADV,i}(t)),$$

Outcomes of interest:

- Time to re-hospitalization (recurrent events)
- Treatment effect
- Tumor necrosis factor (TNF-α) biomarker response of different antigens biomarkers measurements at baseline scheduled at 3, 6, 9, and 12 months
- Previous opportunistic infections (OPI)

Research Questions:

- How strongly correlated is the association between TNF-α biomarker and the risk of death of patients?
- Is TNF-α biomarker response a good biomarker to determine who may have a poor response to **tumor necrosis factor-alpha** inhibitor TB drugs?

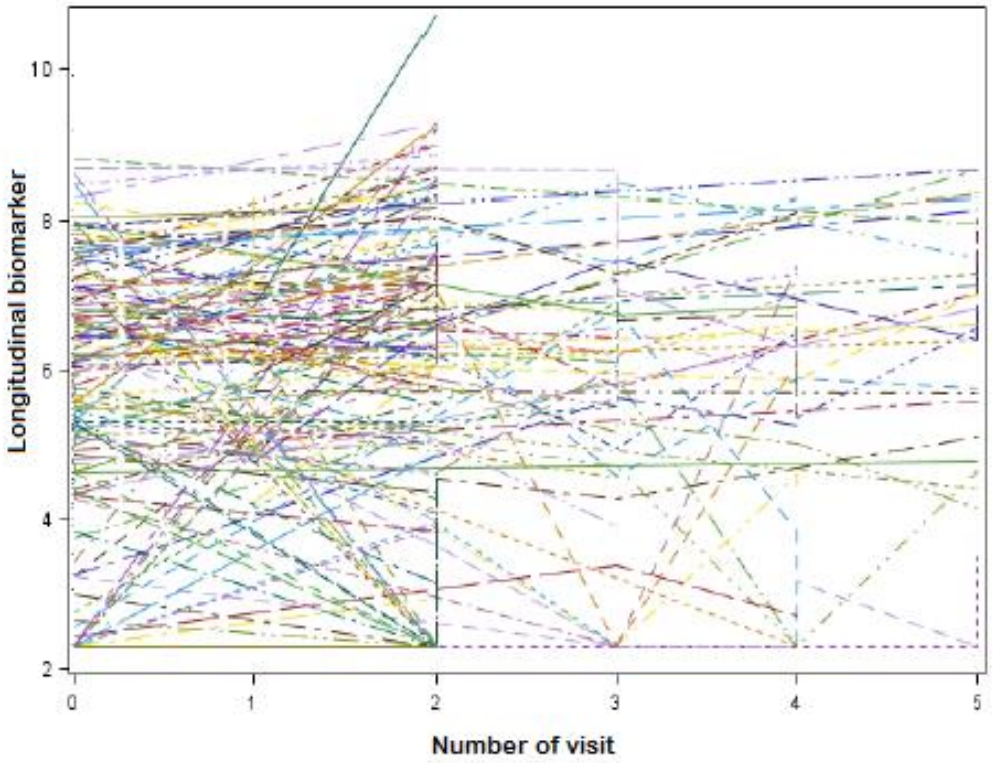
  * if treatment improves TNF-α response, does it also improve survival?

- So far, when we fit the joint model we assumed that the visiting process is non-informative

Motivating study Data

- 475 TB patients who had failed or were intolerant to TB treatment
- The TST was scheduled at baseline, 1, 3, 6, 12 and 24 months or until TST conversion occurred.
- Blood samples for TST negative at baseline scheduled at 1, 3, 6, 9, 12 months, as well as at sick visits.
- The baseline information: age, sex and Purified Protein Derivative reading
- Discrete time periods for conversion define as
  - T1=baseline to month 3 as period 1,
  - T2=3-6 months as period 2,
  - T3=6-12 months as period 3, and
  - T4=12-24 months as period 4.
- There was no lost to follow-up prior to 24 months and Censoring only take place at 24 month
- The conversion times or censoring status as the conversion survival data.
- The TNF-α or IFN-γ response to stimulation of different antigens as longitudinal biomarkers.

Application to a real data



- A spaghetti plot showing individual trajectories of TB biomarker shows that the lines become sparse after the 2nd visit.
- The MTB increases in the first two visits for most of the subjects, then increases slowly or stays flat for most subjects

Risk Estimate with 95% credible interval

Covariates	Two-step Model Relative Risk (95% CI)	P-Value	Joint Model Relative Risk (95% CR)
Sex			
Men	1.0		1.0
Women	0.96 (0.81,1.13)	0.60	0.97 (0.82,1.16)
Smoking			
Nonsmoker	1.0		1.0
Former	1.17 (0.98,1.41)	0.09	1.20 (0.99,1.44)
Current	1.52 (1.19,1.94)	<.001	1.53 (1.21,1.95)
Diabetes			
No	1.0		1.0
Yes	1.66 (1.40,1.97)	<.001	1.69 (1.43,2.00)
HIV			
No	1.0		1.0
Yes	1.39 (1.12,1.72)	0.003	1.35 (1.08,1.66)
BMI	1.08 (1.00,1.17)	0.07	1.08 (1.00,1.17)
TNF-alpha	2.48 (2.00,3.07)	<0.001	4.54 (3.03,5.55)

Covariates	P-Value	Joint Model Relative Risk (95% CR)
TNF-alpha	<.001	3.95 (3.18,4.71)
Average slope of TNF-alpha	<0.001	2.39 (1.40,3.73)

- The estimated relative risk association with time-invariant baseline survival covariates are similar between the two models
- ERR association is much larger under the proposed model. This is expected as our model is capable of estimating subject-specific TB biomarker over time and capable of accurately testing the association between biomarker value if death and risk of death.
- Both model identified the biomarker as a significant risk factor mortality
- In the model, it is estimated that every o.1ml of 5 tuberculin units of purified protein derivative is associated with 4.5times higher risk of death.

Model II:

- we adjusted for the area under the derivative curve of TB biomarker from the time the followup starts until the survival time (time of death/censored)
- every o.1ml of 5 tuberculin units of purified protein derivative is associated with 3.95times higher risk of death.
- Higher average slope of biomarker is associated with 2.3times higher risk of death.

Performance Selection for the two model

Model I	Brier Score	H-L	H-L Test P-Value	K-S	K-S Test P-Value
Joint	0.1302	28.3484	0.0122	0.5156	0.0000
Twostep	0.1921	8.3711	0.4700	0.1408	0.0789

Model II	AUC	Brier Score	H-L	H-L Test P-Value	K-S	K-S Test P-Value
Joint	0.8780	0.1302	28.3484	0.0122	0.5156	0.0000
Twostep	0.7089	0.1723	7.9086	0.5008	0.3141	0.0000

Models	AUC Difference	Mean	tValue	DF	P-Value
1	Joint	0.434	336.076	892	0.0000
	Twostep	0.299	273.672	892	0.0000
2	Joint	0.327	113.583	892	0.0000
	Twostep	0.188	111.349	892	0.0000

- The model with smaller Brier score is preferred
- Larger value for AUC, Hosmer-Lemeshow (H-L) and Kolmogorov-Smirnor (K-S) are desirable for a model prediction

- The joint model is significantly better with mean diff. higher

Conclusion, Limitation and future work

- The proposed Dirichlet process mixture models to model longitudinal data with latent sub-group random intercepts compared to the common longitudinal models, perform the best in terms of the Brier score accuracy probabilistic prediction of estimating conditional covariate effects.
- Sensitivity analyses to the robustness of the model.
- The survival component of our model still relies on the proportional hazard assumption.
- In future, our modeling framework can be extended to include a more general non-proportional hazard survival models that can also include time-dependent coefficients inside the survival model.

Thanks for listening