KWARA STATE UNIVERSITY, MALETE

School of Postgraduate Studies (SPGS)

BAYESIAN PENALIZED CREDIBLE REGION IN HIGH DIMENSIONAL SURVIVAL DATA

BY

AJIBADE KUNLE SAHEED

19/57MST/00002

January, 2022

KWARA STATE UNIVERSITY, MALETE

SCHOOL OF POSTGRADUATE STUDIES (SPGS)

BAYESIAN PENALIZED CREDIBLE REGION IN HIGH DIMENSIONAL SURVIVAL DATA

M.Sc. THESSIS SUBMITTED AND PRESENTED

BY

AJIBADE KUNLE SAHEED

19/57MST/00002

In Partial Fulfillment of the Requirements for the Award of Masters of Science, Degree in Statistics,

DEPARTMENT OF STATISTICS AND MATHEMATICAL SCIENCES,

FACULTY OF PURE AND APPLIED SCIENCES,

KWARA STATE UNIVERSITY, MALETE,

NIGERIA.

January, 2022

DECLARATION

I hereby declare that this research project titled **"BAYESIAN PENALIZED CREDIBLE REGION IN HIGH DIMENSIONAL SURVIVAL DATA"** is my own work and has not been submitted by any other person for any degree or qualification at any higher institution. I also declare that the information provided therein are mine and those that are not mine are properly acknowledged.

Ajibade Kunle Saheed — Signature and Date

APPROVAL PAGE

This is to certify that this thesis by (AJIBADE KUNLE SAHEED) has been read and approved as meeting the requirements of the Department of mathematics and statistics, Kwara state University, Malete For the award of the degree of masters (M.Sc) in Statistics.

.....	.....
Dr. K. A. Dauda	Date
Supervisor	
.....	.....
Prof. W.B Yahaya	Date
Co-Supervisor	
.....	.....
Prof. A. AbdulRaheem	Date
Head of Department	
.....	.....
Internal Examiner	Date
.....	.....
External Examiner	Date
.....	.....
Prof. Hamza I. AbdulRaheem	Date
Dean, School of Postgraduate Studies (SPGS)	

DEDICATION

This project is dedicated to the Almighty Allah.

ACKNOWLEDGMENTS

My sincere gratitude goes to ALLAH who made this movement successful and also to my supervisor; Dr. K.A Dauda, a mentor, in gratitude, your immense contributions to this study will never be forgotten.

I would also like to thank the Head of Department; Dr. AbdulRaheem AbdulRazaq and all lecturers of the Department of Mathematics and Statistics, for their coordinating supports.

Finally, to share in the joy and success are my parents: Mr M. O. Ajibade & Mrs R. T. Ajibade, my siblings: Amirah, Monsur, Monshood, Mubaraq, Fathiah Muhideen, Zemirah and my dear Kadijah for their consistent support throughout my time in KWASU.

Ajibade S.K.

October, 2021

TABLE OF CONTENTS

List of Tables

Abstract

The technique of Bayesian process in selecting variable using credible regions penalized, perform model fitting and variable selection in survial analysis via cox's proportional hazard model. The intrest is to extract the sparsed solution within the joint posterior credible regions. More recently, advancements in the intrest of global-local shrinkage priors have been made for high-dimensional Bayesian variable selection. However, no much intrest of studies have been done to incorporate global-local shrinkage priors called Dirichlet-Laplace (DL) prior. This study is motivated to incorporate and adapt global-local shrinkage priors (DL prior) to cox's proportional hazard (Cox's PH) model and the tunnig of hyperparameters in prior distributions for cox's PH model. The adapted method was compared with the existing method using AIC, BIC and DIC. The result over the simulation and real life data revealed that the modified model outperforms the existing method. Additionally, the proposed modified model selected minimum numbers of genes than the existing method.

Keywords: Variable selection, Cox's PH, Global-local shrinkage prior, DIC, Hyperparameters, DirichletLaplace

CHAPTER ONE

GENERAL INTRODUCTION

1.1 Background of Study

Survival analysis characterizes relationships between time-to-event endpoints and multiple explanatory variables (covariates) (Dauda KA, et al.,2019, Kalbfleisch and Prentice, 1980; Oakes, 2001). For instance how much time will take for an event to happen, like

- In Medical, time until tumor reoccure or time until Acquired Immunodeficiency Syndrome (AIDS) for Human Immunodeficiency Virus (HIV) patients.
- In Engneering, time until a machine part fails or time until a lighting bulb fault.
- In Social Science, time until a graduate from teachers college got a job.

From the literature, a well known approach from frequentist to the above problems are non-parametric model, semi-parametric model and parametric model.

Nonetheless, in many applications, it is paramount to obtain a accurate or practicable characterization of uncertainty in estimates of parameters, functions of parameters, and predictions. Usual frequentist approaches to characterize uncertainty, for instance, constructin of asymptotic confidence regions or using the bootstrap, may not work in high-dimensional settings (Fan and Lv 2004, Tibshirani 1996). For instance, in cox's regression when the number of sample size n is equal to or lesser than the number of

independent variable p, one cannot naively appeal to asymptotic normality and an adequate characterization of uncertainty may not be provided through resampling procedure from the data. Most penalized estimators correspond to the mode of a Bayesian posterior distribution. For example, Dirichlet-Laplace global-local shrinkage priors (Yan and Howard, 2016) estimation under a normal linear regression model having a Laplace kernel (DE(b)) with Dirichlet distribution and gamma (pa,1/2) prior on the coefficients. Given this connection, a Bayesian process has vivid advantages in terms of tuning parameter choice, permitting crucial penalty parameters to be marginalized over the posterior distribution instead of relying on cross-validation.

1.2 Statement of Problem

The staggering emphasis in the literature on high-dimensional time to event data analysis has been on speedy generating point estimates with satisfied empirical and theoretical properties. In this study, we extended the global-local shrinkage prior to cox's proportional hazards model using beyesian approaches. The intrest is to search for the sparsed model within the joint posterior credible regions

1.3 Aim and Objectives of the study

Aim

In the cause of this study, we proposed a Dirichlet Global-local shrinkage priors that will spars high dimensional time to event data and predict effectiflly.

Objectives

- To obtain the posterior for high dimensional survial data via global local shrinkage prior.
- To show that proir used is in order.
- Shrinking of small coefficent greatly tight to zero and rearly shrink large coefficients.

1.4 Chapter Outline

This research thesis is organized as follows. Chapter 1 covers the general introduction. Chapter 2 reviews the Cox's PH models and penalized credible region variable selection method. Chapter 3 details the proposed method which combines likelihood, shrinkage priors and penalized credible region variable selection. Chapter 4 reports the simulation results and real life data result, and Chapter 5 gives the summary of the research findings and conclusion.

1.5 Significance of the study (Global Local Shrinkage prior)

Having a high dimensional data i.e $n \gg p$, there are many approaches or methods of feature selection and Bayesian approach stated in the literature review, but our proposed method;

- Makes use of the foredeal of global-local (GL) shrinkage priors, which can effectively shrink small coeffcients and reliably estimate the coeffcients of important variables simultaneously.
- The motivation is that concentrated posteriors is achived using GL shrinkage priors within the same credible region level, thus having good performance for variable selection, by finding sparsed solutions more easily (Yan and Howard, 2016).

CHAPTER TWO

REVIEW OF COX'S AND SOME PENALIZED GLOBAL-LOCAL SHRINKAGE PRIOR

2.1 Bayesian Perspective towards Survival Analysis

2.1.1 Parametric Modeling Approach

Parametric modeling is straightforward, and many Bayesian analyses in practice are based on a parametric model. Zhenyu (2014) express the Weibull model, is one of the most extensively used parametric survival model combined with gamma, to produce the combined posterior of β which does not have a closed form, so numerical integration or MCMC methods are used to estimate the posterior distribution of β.

2.1.2 Semiparametric Modeling Approach

Many literature have consided Bayesian semiparametric approach for the accelerated failure time model and Bayesian Cox proportional model (Zhenyu 2014).

2.1.3 Cox's Proportional Hazards models

A Cox proportional hazards (PH) model is accepted mathematical model used for modelling survival data. This model was proposed by Cox and Oakes (1972) and has also known as the Cox regression model. The reason why the Cox PH model is so demanding in survival analysis is because it is a semiparametric and a "robust" model.

Let T , C, and X denote the survival time, the censoring time, and their associated covariates, respectively. Y is derived from min { T, C } which is observed time and $\delta = \mathtt{I}(\mathrm{T} \leq \mathrm{C})$ the censoring indicator. We assume that T and C are conditionally independent given X and the censoring mechanism gives no information. Our sampled data set $\{(x_i, y_i, \delta_i) : x_i \epsilon \mathbb{R}^p, y_i \epsilon \mathbb{R}^+, \delta_i \epsilon \{0,1\}, i = 1, 2, \cdots, n\}$ is an independently and identically distributed random sample from a certain population (X, Y, δ). Define $C = \{i : \delta_i = 0\}$ and $U = \{i : \delta_i = 1\}$ to be the censored and uncensored index sets, respectively.

$$h(t|x) = h_o(t) exp^{x^T \beta} \tag{2.1}$$

where $h_o(t)$ is the baseline hazard function. In this model, both $h_o(t)$ and β are unknown and have to be estimated. Under model (2.1), Then the complete likelihood of the observed data set is given by

$$L = \prod_{i \epsilon U} f(t_i|x_i) \prod_{i \epsilon C} \bar{F}(t_i|x_i) = \prod_{i \epsilon U} h(t_i|x_i) \prod_{i=1}^{n} \bar{F}(t_i|x_i)$$

where $f(t|x), \bar{F}(t|x) = \int_t^\infty f(s|x)ds$, and $h(t|x) = f(t|x)/\bar{F}(t|x)$ are the conditional density function, the conditional survival function, and the conditional hazard function of T given X = x, respectively, see J. Fan,(2010).

$$L = \prod_{j=1}^{N} h_o(t_{(j)}) exp^{x_{(j)}^T \beta} \prod_{i=1}^{n} exp - H_o(t_i) exp(x_{(i)}^T \beta) \tag{2.2}$$

Following Breslow's idea, equation (2.2) becomes, see (Fan et al 2010)

$$H_0(t_i) = \sum_{j=1}^{N} h_j I(i \epsilon \mathbb{R}(t_j^0)) \tag{2.3}$$

Result of the log-likelihood becomes

$$log(L) = \sum_{j=1}^{N} \left\{ log(h_j) + x_{(j)}^T \beta \right\} - \sum_{i=1}^{n} \left\{ \sum_{j=1}^{N} h_j I(i \epsilon \mathbb{R}(t_j^0)) exp(x_{(i)}^T \beta) \right\}.$$

Maximizer h_j is given by

$$\hat{h}_j(\beta) = \left\{ \sum_{i \epsilon \mathbb{R}(t_j^0)} exp(x_{(i)}^T \beta) \right\}^{-1} \tag{2.4}$$

Putting this maximizer back to the log-likelihood, we get

$$log(L) = \sum_{i=1}^{n} \left\{ \delta_i x_{(i)}^T \beta - \sum_{i=1}^{n} \delta_i log \left\{ \sum_{j \epsilon \mathbb{R}(y_i)} exp(x_{(j)}^T \beta) \right\} \right\}$$

Adding the censoring indicatorδ_i, which is equivalent to

$$L(\beta) = \sum_{j=1}^{N} \left\{ x_{(j)}^T \beta - log \left\{ \sum_{i \epsilon \mathbb{R}(t_j^0)} exp(x_{(i)}^T \beta) \right\} \right\} \tag{2.5}$$

2.2 Penalized Methods

literature has shown the basic penalized methods for regression which are: The Least Absolute Shrinkage and Selection Operator (Lasso; Tibshirani 1996), smoothly clipped absolute deviation (SCAD; Fan & Li 2001), the elastic net (Zou & Hastie 2005), adaptive Lasso (Zou 2006), the Dantzig selector (Candes & Tao 2007), and octagonal shrinkage and clustering algorithm for regression (OSCAR; Bondell & Reich 2008). Zhang et al. (2018) proposed a Dirichlet process prior on the accelerated failure time (AFT) model. In addition, following the data augmentation approach of Tanner and Wong (1984), the censored observations can be imputed (Bonato et al., 2011).

When developing the penalized Cox's model is of particular interest, then there exist several proposals in the literature; for example, Sushil et al.(2013) derived High-dimensional, massive sample-size Cox proportional hazards regression; Fan et al.(2010) also derived a penalization for cox's model via Sure Independence Screening (SIS) and Iterative sure independence screening (ISIS). However, they did not consider Bayesian settings in their research. (Bhattacharyya et al., 2021) comes up with the procedure to detect best variable selection and future prediction for classification problems under survival analysis using Bayesian procedure embedded with Global-Local Shrinkage Priors.

But interest of this work does not solve classification (categorical response or pridictors) problem. However (Shimamura et al., 2021) developed clustering (unsupervised learning) method with Bayesian process based on the ideas of lasso and global local shrinkage priors.

2.2.1 Bayesian framework

In the Bayesian framework, traditional Bayesian approaches for variable selection is either relying on the calculation of posterior inclusion probabilities for each predictor or each possible model, or a choice of posterior threshold, e.g Stochastic search variable selection (SSVS) (George & McCulloch 1993), Bayesian regularization (Park & Casella 2008, Li et al. 2010, Polson et al. 2013, Leng et al. 2014), Empirical Bayes variable selection (George & Foster 2000), Spike and slab variable selection (Ishwaran & Rao 2005). Global-local (GL) shrinkage priors. i.e "global" term should provide substantial shrinkage towards zero, "local" terms should have heavy tails so that "signals" are not shrunk too much.
Different type of shrinkage priors for β are as follow

- The horseshoe prior (Carvalho, Polson and Scott, 2010, Biometrika).
- The horseshoe+ prior (Bhadra et al., 2016, Bayesian Anal.).
- The hypergeometric inverted-beta prior (Polson and Scott, 2010, Bayesian Anal.).
- The generalized double Pareto prior (Armagan, Dunson and Lee,2013, Stat. Sinica).
- The three parameter beta prior (Armagan, Dunson and Clyde, 2011,NIPS).
- The Dirichlet-Laplace prior (Bhattacharya et al., 2015, JASA).

In all these papers, the Maity et al 2019 is the similar work to the proposed study. How ever, Maity et al 2019 focused on Horseshoe prior for accelerated failure time AFT (log normal) and our proposed Dirichlet laplas prior focused on semiparametric modeling approach (Cox's PH), see the figure below.

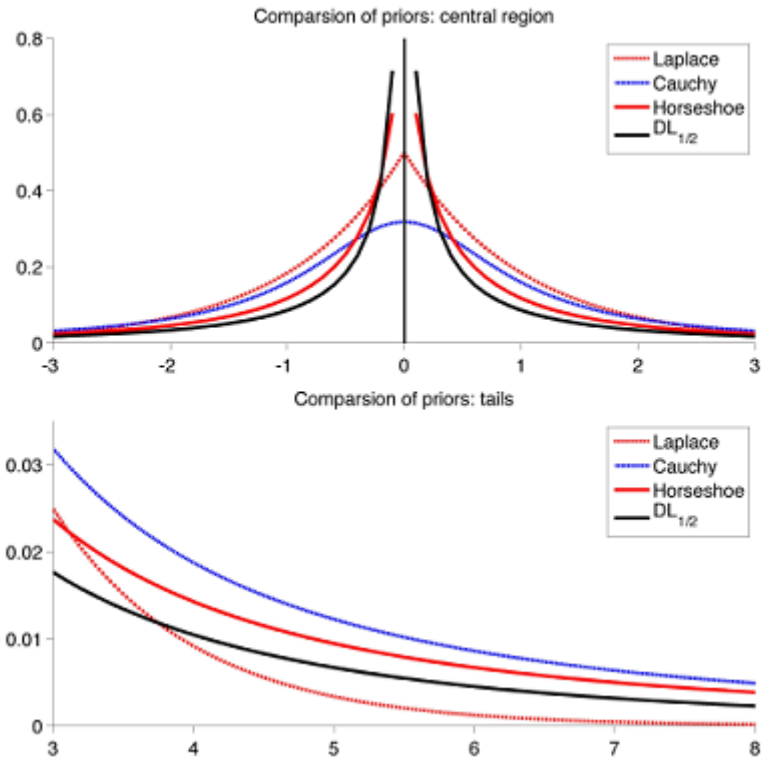

Figure 2.1: Comparison of Priors based on the Central Region and Tails

CHAPTER THREE

METHODOLOGY

The posterior distribution is the most important quantity in Bayesian method. It contains all the information available about the unknow parameter θ after having observed the data $X = x$, where θ represents the prior information and data $X = x$ represents the likelihood function.

$$\begin{gathered} Posterior\ Distribution = Prior\ Distribution\ \times\ Likelihood\ Function \\ f(\theta|D)\ =\ f(\theta)\ \times\ f(D|\theta) \end{gathered} \tag{3.1}$$

Since we adopt the existing Cox's proportional hazards (PH) model as our likelihood with two unknow parameters, the partial likelihood is due to Cox (1975). Maximizing $L(\beta)$ in (2.5) with respect to β, we can get an estimate $\hat{\beta}$ of the regression parameter. Once $\hat{\beta}$ is available, we may plug it into (2.4) to get $\hat{h}_j(\hat{\beta})$. These newly obtained $\hat{h}_j(\hat{\beta})$s can be plugged into (2.3) to obtain our nonparametric estimate of the baseline cumulative hazard function.

3.1 Dirichlet-Laplace Priors

Survival data are rarely Normally distributed, but are skewed and comprise typically of many early events and relatively few late ones (Clark et al 2003). For the normal mean model, Bhattacharya et al. (2015) proposed a new class of DirichletLaplace (DL) shrinkage priors, possessing the optimal posterior concentration property. We construct the

generalization of the DL priors for the Cox's regression model. The proposed hierarchical DL prior is as follows: for $j = 1, \cdots, p$,

$$\begin{aligned} \beta_j|\sigma, \phi_j, \tau &\sim DE(\sigma, \phi_j, \tau), \\ (\phi_1, \ldots, \phi_p) &\sim Dir(s, \ldots, s), \\ \tau &\sim Ga(pa, 1/2). \end{aligned} \tag{3.2}$$

where DE(q) indicate a zero mean Laplace kernel with density $f(y) = (2q)^{-1} \exp\{-|y|/q/\}$ for $y\epsilon\mathbb{R}^+$, Dir$(s, \ldots, s)$ is the Dirichlet distribution with concentration vector (s,..., s), and $Ga(pa; 1/2)$ indicate a Gamma distribution with shape pa and rate 1/2. In (3.2), small values of s would shrink most of $(\phi_1, \ldots, \phi_p)$ to be close to zero and only the most important remain as nonzero; while large values allow less singularity at zero, thus controlling the sparsity of regression coeffcients. The ϕ_j's are the local scales, allowing deviations in the degree of shrinkage (Yan et al 2016). As spoted out in Bhattacharya et al., (2015), τ controls global shrinkage towards the origin and to some extent determines the tail behaviors of the marginal distribution of β_j's.

3.2 Posteriors Computation

Bayesian computation generally exploits modern computer power to carry out simulations based on Markov chains and is known as Markov chain Monte Carlo (MCMC) (Spiegelhalter et al., 2004). Monte Carlo methods are techniques that have the aim of evaluating integrals rather than exact or approximate algebraic analysis (Spiegelhalter et al., 2004). Several MCMC algorithms that are commonly used are Gibbs sampling, Metropolis-Hastings, reversible jump, slice sampling, particle filters, perfect sampling and adaptive rejection sampling.

Due to computational complexity of the posterior, we adopt gibbs sampling technique which is one of the Monte Carlo Markov Chain (MCMC) procedure, we modified gibbs sampling steps to accommodate the Cox's regression model.

$$f(\psi, \phi, \tau|\beta, \sigma^2) = f(\psi|\phi, \tau, \beta, \sigma^2) \times f(\tau|\phi, \beta, \sigma^2) \times f(\phi|\beta, \sigma^2) \times (\psi\phi\tau). \tag{3.3}$$

The equation (3.3) is independent of $y \equiv h(t|X)$ conditionally on β and σ^2.
These are the following procedure to generate our posterior parameters and summary of the Gibbs sampling steps:

- Sample $\sigma^2|\beta, \psi, \phi, \tau, y$. We draw σ^2 from an inverse Gamma distribution, $IG(a_1 + (n+p)/2, b_1+(\beta^T S^{-1}\beta+(Y-X\beta)^T(Y-X\beta))/2)$, where $S = diag(\psi_1\phi_1^2\tau^2, \cdots, \psi_p\phi_p^2\tau^2)$.
- Sample $\beta|\psi, \phi, \tau, \sigma^2, y$. We draw β from $N(\mu, \sigma^2 V)$, where $V = (X^T X + S^{-1})^{-1}$ with the same S as above, and $\mu = VX^T Y = (X^T X + S^{-1})^{-1}(X^T Y)$.
- (iii) Sample $\psi|\phi, \tau, \beta, \sigma^2$, we first draw $\psi_j^{-1}|\phi_j, \tau, \beta, \sigma^2, j = 1, \cdots, p$, independently from the distribution InvGaussian $(\mu_j = \sigma\phi_j\tau|\beta_j|, \lambda_0 = 1)$, where InvGaussian (μ, λ_0) denotes the inverse Gaussian with density $f(y) = \sqrt{\lambda_0/(2\pi y^3)}exp\{-\lambda_0(y-\mu)^2/(2\mu^2 y)\}$ for $y > 0$. Then take the reciprocal to get the draws of $\psi_j (j = 1, \cdots, p)$.
- Sample $\tau|\phi, \beta, \sigma^2$, We draw τ from a $giG(\chi = 2\Sigma_{j=1}^p|\beta_j|/(\phi_j\sigma),\ \rho = 1,\ \lambda_0 = pa - p)$.
- Sample $\phi|\beta, \sigma^2$. We draw $T_1, \cdots, T_p$ independently with $T_j \sim giG(\chi = 2|\beta_j|/\sigma, \rho = 1, \lambda_0 = a - 1)$, then set $\phi_j = T_j/T$ where $T = \Sigma_{j=1}^p T_j$.

3.3 Model Evaluation and Criteria

In this case, model criteria will be more informative. We consider three popular model choice criteria: the deviance information criterion (DIC)(Spiegelhalter et al., 2002), which is hierarchical modeling generalization of Akaike information criterion (AIC). Wildely used in Bayesian model selection problems where the posterior distributions of the models have been obtained by Markov chain Monte Carlo (MCMC) simulation. Akaike information criterion (AIC)(Akaike 1998) and Bayesian information criteria (BIC) (Schwarz et al. 1978)

$$DIC = -\frac{4}{R}\sum_{r=1}^{R} log\{\prod_{i=1}^{n} L(D_i|\gamma)\} + 2log\{\prod_{i=1}^{n}\frac{1}{R}\sum_{r=1}^{R} L(D_i|\gamma)\} \tag{3.4}$$

$$AIC = 2k - 2ln(L(D_i|\gamma)) \tag{3.5}$$

$$BIC = kln(n) - 2ln(L(D_i|\gamma)) \tag{3.6}$$

where $L(D_i|\gamma) = f(\psi, \phi, \tau|\beta, \sigma^2)$.

n = the number of observations, or equivalently, the sample size.

k = the number of parameters estimated by the model.

3.4 Algorithm

Algorithm 1 Cox's with Dirichlet-Laplace shrinkage prior

input Data set of time to event $(y_i, x_{ij}) : y_i \epsilon Y$ and Y is a function of time (T) but not involve the covarate $x_{ij}, i = 1, 2, \cdots, n,\ j = 1, 2, \cdots, p$

output return the most important selected variable

```
for i = 1 to nmc number of posterior draws to be saved do
    if x_ij s contain latent variable then
        the data is string and can not be processed
    else
        x_ij s is multivariate normal.
        Extract the best turning hyperparameter
    end if
end for
if turning hyperparameter is greater than 1/p then
    compute 1/maximum(n,p) as the value of hyperparameter
else
    Sample all the model parameters and proceed to Monte Carlo Markov Chain
    (MCMC) procedure using gibbs sampling technique
end if
```

The Algorith (1) above take in high dimentional survival (time to event) data as input, process it through methodology explaned above and retun the best and most important predictors variable selected as an output.

CHAPTER FOUR

RESULT AND DISCUSSION

4.1 Simulation Scheme

We simulated scenarios data to illustrate the methodologies discussed in this study. For the simulation, we considered multivariate normal predictors variable ($X = 100$), also a reponse $y = 100$ were drawn from normal distribution and find it's exponential to have a skewed right response as time(T) and censoring time were drawn from gamma distribution to determine the status C. We repeated the procedure for

$n = 100$ & $X = 200$,

$n = 100$ & $X = 500$,

$n = 100$ & $X = 1000$,

$n = 100$ & $X = 2000$,

$n = 100$ & $X = 5000$.

4.2 Result from Simulation Scheme

The table 4.1 below shows the result of simulation scheme through the increases of numbers of predictors (p) 100, 200, 500, 1000, 2000 and 5000. We compared our method "DL priors" with another Global local shrinkage prior which is "Horseshoe prior", we evaluated the procedure with deviance information criterion (DIC), Akaike information criterion (AIC) and Bayesian information criteria (BIC).

Table 4.1: Result of AIC, BIC and DIC on Horseshoe and DL prior

p	GL prior	AIC	BIC	DIC
100	Horseshoe	8.8682	20.7680	7.5986
	DL prior	7.3386	20.3645	6.6782
200	Horseshoe	8.6882	19.6721	7.2837
	DL prior	7.4532	17.5973	6.4591
500	Horseshoe	9.2842	17.2617	5.9244
	DL prior	8.7689	15.9573	4.8670
1000	Horseshoe	9.5782	16.6232	4.1098
	DL prior	8.6725	14.3451	3.2702
2000	Horseshoe	9.4326	15.6322	3.6773
	DL prior	8.1746	13.9756	3.1022
5000	Horseshoe	9.7283	15.2203	3.4234
	DL prior	8.1943	12.8572	2.9786

Discussion on the table 4.1. For the above priors compared, using the model criteria that "the lower the AIC, BIC and DIC value the better the model" going through diffrent level of simulation scheme, at constant sample size of 100 at all levels, when number of predictors is 100, 200, 500, 1000, 2000 and 5000 (DL prior) preformed better over the Horseshoe prior across all the criteria used and down the diffrent leavels of predictors (p).

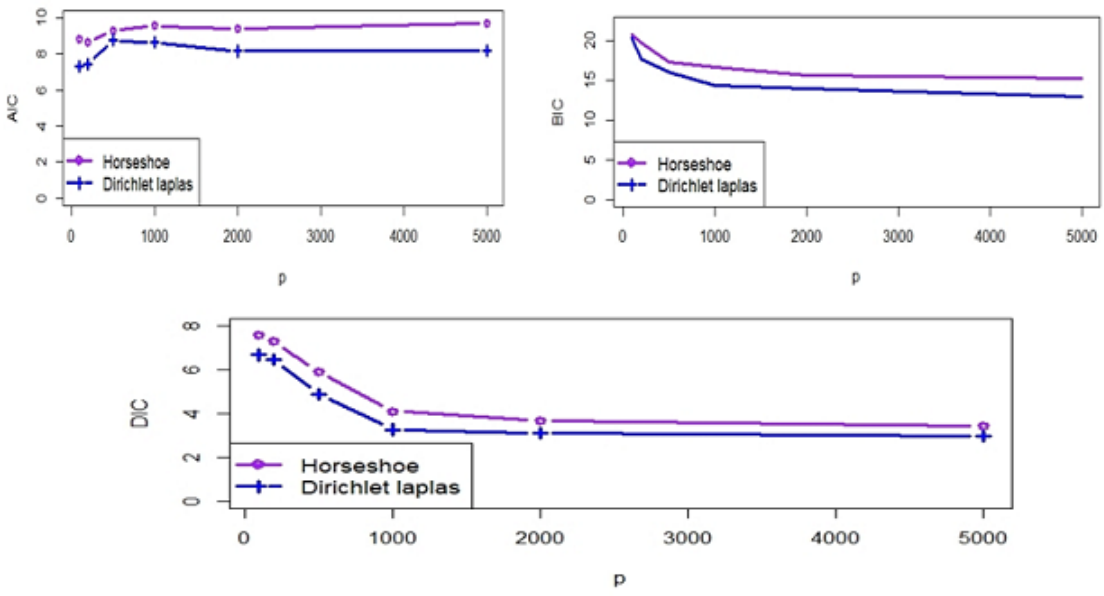

Figure 4.1: Graphical representation of AIC, BIC and DIC on Horseshoe and DL priors

From figure 4.1, shows the plots of AIC, BIC and DIC value aganst the number of predictors (p) on three diffrent plot. From all plot DL proir with blue colour on the graph shows the minimum value of AIC, BIC and DIC

4.3 Real Life Data application

The data was sourced from R package **"bujar"** version 0.2, 2009. Microarray data of DLBCL of 181 patients treated with a combination chemotherapy with cyclophosphamide, doxorubicin, vincristine and prednisone (CHOP). The first column is the survival times. The second column is an indicator whether the survival time was observed or right censoring occurred. 0=alive, 1=dead. There are 3833 genes after the filtering process.

Table 4.2: Data Chop Discrption

S/N	survtime	status	X1552325 at	X1552365 at	$\cdots$, 3833
1	2.68	1	4.982	2.322	$\cdots$,
2	0.82	1	0.926	4.329	$\cdots$,
3	2.54	1	4.863	1.848	$\cdots$,
4	9.67	0	3.776	2.104	$\cdots$,
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
181			$\leftarrow\leftarrow$		$\Leftarrow$

Table 4.3: variable selection for priors on Real life Data "CHOP"

prior	Number of gens selected
Horseshoe	27
(DL) shrinkage prior	13

After carefull implimentation of the modified method, proposed Dirichlet-Laplace(DL) shrinkage prior with Cos's PH selected minimum numbers of genes (13) than the existing Horseshoe prior method (27), this were revealed in the table (4.3) for the purpose of variable selection.

Furthermore, table (4.4) showcased the model accuracy through the AIC, BIC and DIC. We also observed that the DL shrinkage prior outperformed the existing method respectively.

Table 4.4: Result of AIC, BIC and DIC for Horseshoe and DL prior on Real life Data "CHOP"

prior	AIC	BIC	DIC
Horseshoe	5.87254	19.75894	4.58794
(DL) shrinkage prior	2.89197	18.88446	0.98472

Table 4.5: Selcted genes for Horseshoe and DL prior on "CHOP"

DL prior		Horseshoe prior			
X1557431 at	X1558999 x at	X1552381 at	X1558999 x at	X1552390 a at	X1552325 at
X1560884 at	X1563845 at	X1552439 s at	X1565162 s at	X1560884at	X1552389 at
X1565162 s at	X207166 at	X1552373 s at	X1552365 at	X1552368 at	X1552394 a at
X220118 at	X224396 s at	X1565162 s at	X224396 s at	X207166 at	X1552372 at
X229839 at	X234849 at	X1552367 a at	X1557431 at	X1552393 at	X1563845 at
X240657 at	X244546 at	X1552430 at	X1553587 at at	X1552514 at	X244682 at
X244682 at		X244824 at	X1553547 at	X1552514 at	

The names and the type of selected genes were presented in the the table (4.5) for both the proposed and the existing methods. Nine genes were selected by both methods.

CHAPTER FIVE

CONCLUSION

5.1 Conclusions

We have described a new Bayesian framework for Cox's proportional hazards models where the cumulative baseline hazard function was modeled with a Dirichlet-Laplace (DL) shrinkage prior. The proposed approach helps us to select the minimum and most important variables so as to avoid the non informative variables in the survival dataset. When compared to the model with the horsehoe prior, the proposed DL shrinkage prior characterized to have selected the minimum number of predictor variables with a good fit at all levels of number of pridictor variations.

5.2 Limitations of Study

In this study, we have only considered the Cox's proportional hazards model under semi parametric model, added with a Dirichlet-Laplace (DL) shrinkage prior through MCMC scheme. Nonetheless, the extension to other survival distributions (weibul, exponential, log-normal etc) is immediate. Furthermore, the use of the latent variable technique may also make possible extension in the further studies.

References

[1] Akaike, H. (1998), Information theory and an extension of the maximum likelihood principle, in 'Selected Papers of Hirotugu Akaike', Springer, pp. 199-213.

[2] Bhattacharyya1 A., Pal1 S., Riten M. and Shesh N. Rai (2021). Applications of Bayesian Shrinkage prior models in clinical research with categorical responses, Research Square. DOI: https://doi.org/10.21203/rs.3.rs-137866/v1

[3] Bhattacharya, A., Pati, D., Pillai, N. S. & Dunson, D. B. (2015), Dirichlet Laplace priors for optimal shrinkage, Journal of the American Statistical Association 110(512), 1479 1490.

[4] Cox, D. R. (1975). Partial likelihood. Biometrika 62 269-76.

[5] Ishwaran, H. & Rao, J. S. (2005), 'Spike and slab variable selection: frequentist and bayesian strategies', Annals of Statistics pp. 730-773.

[6] J. Fan and H. Peng, "Nonconcave penalized likelihood with a diverging number of parameters," The Annals of Statistics, vol. 32, no. 3, pp. 928-961, 2004.

[7] K. Shimamura1, S. Kawano. Bayesian sparse convex clustering via global-local shrinkage priors, Computational Statistics (2021) 36:2671-2699 https://doi.org/10.1007/s00180-021-01101-7

[8] K.A Dauda, B. Pradhan, B. Uma Shankar, S. Mitra. Decision tree for modeling survival data with competing risks. Biocybern Biomed Eng (2019), https://doi.org/10.1016/j.bbe.2019.05.001

[9] kalbfleisch,j.d.and prentice,r.l.(1980).the statistical analysis of failure timedata. new york: john wiley and sons

[10] Khan, M.H.R. and Shaw, J.E.H. (2016) Variable selection for survival data with a class of adaptive elastic net techniques. Statistics and Computing, 26(3), 725-741.

[11] Khan, M.H.R. and Shaw, J.E.H. (2017) Variable selection for accelerated life-time models with synthesized estimation tech-niques.Statistical Methods in Medical Research,1-17.

[12] Lenz, et al. (2008). Stromal gene signatures in large-B-cell lymphomas. New England Journal of Medicine, 359(22), 2313-2323

[13] Liu, B., Li, Y., Sun, Z., Ghosh, S. and Ng, K. (2018, February)Early Prediction of Diabetes Complications from Electronic Health Records: A Multi-Task Survival Analysis Approach. Thirty Second AAAI Conference on Artificial Intelligence (AAAI-18), New Orleans, LA.

[14] Maity AK, Bhattacharya A, Mallick BK, Baladandayuthapani V. Bayesian data integration and variable selection for pancancer survival prediction using protein expresion data. Biometrics.2019;1-10.

[15] Oakes,D.(2001). Biometrika centenary: survival analysis.Biometrika88(1), 99-142.

[16] T G Clark, M J Bradburn, S B Love & D G Altman. Survival Analysis Part I: Basic concepts and first analyses. British Journal of Cancer volume 89, pages232-238 (2003)

[17] R version 4.0.3 (2020-10-10) "Bunny-Wunnies Freak Out" Copyright (C) 2020 The R Foundation for Statistical Computing. Platform: x86 64-w64-mingw32/x64 (64-bit)

[18] Spiegelhalter, D. J., Abrams, K. R., & Myles, J. P. (2004). Bayesian approaches to clinical trials and health-care evaluation (Vol. 13). Wiley. com.

[19] Schwarz, Gideon E. (1978), "Estimating the dimension of a model", Annals of Statistics, 6 (2): 461-464, doi:10.1214/aos/1176344136

[20] Wang, L., Li, Y., Zhou, J., Zhu, D. and Ye, J. (2017) Multi-task survival analysis.2017 IEEE International Conference on Data Mining (ICDM), IEEE, pp. 485-494.

[21] Yan Zhang & Howard D. Bondell (2016), Variable selection via penalized credible regions with Dirichlet-Laplace global-local shrinkage priors

[22] Zou, H. & Hastie, T. (2005), 'Regularization and variable selection via the elastic net', Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67(2), 301-320.

Appendix

To extract Hazard Function from Time and Status

```
calcna = function(time,event) {
  na.fit = survfit(coxph(Surv(time,event)~1), type="aalen")
  jumps = c(0, na.fit$time, max(time))
  # need to be careful at the beginning and end
  surv = c(1, na.fit$surv, na.fit$surv[length(na.fit$surv)])

  # apply appropriate transformation
  neglogsurv = -log(surv)

  # create placeholder of correct length
  naest = numeric(length(time))
  for (i in 2:length(jumps)) {
    naest[which(time>=jumps[i-1] & time<=jumps[i])] =
      neglogsurv[i-1]   # snag the appropriate value
  }
  return(naest)
}
```

Hyperparameter detection in the prior distribtuion

```
dlhyper<-function(x,y){
  p=ncol(x)
  n=nrow(x)
  #calculate hyperparameter
  xtx=t(x)%*%x
  d=eigen(xtx/n)$values
  P=sum(d)
  Q=4*sum(d^2)-sum(d)^2
```

```
  R=-sum(d)^3
  C=P^2/9-Q/3
  A=P*Q/6-P^3/27-R/2
  B=A^2-C^3
  hyper=sqrt(2/((A+sqrt(B))^(1/3)+sign(A-sqrt(B))*abs(A-sqrt(B))^(1/3)-P/3))
  hyper[hyper<1/p]=1/p
  return(hyper)
}
```

Dirichlet Laplace shrinkage prior in Bayesian Cox's PH regression

```
dl<-function(x,y,burn=5000,nmc=5000,thin=1,hyper=hyper){
  p=ncol(x)
  n=nrow(x)
  #calculate hyperparameter
  xtx=t(x)%*%x
  #initial parameters
  a=rep(hyper,p)
  psi=stats::rexp(p,rate=1/2)
  psi1=rep(0,p)
  phi=LaplacesDemon::rdirichlet(n=1,alpha=a)
  phi[phi <= (1e-40)]<-(1e-40)
  tau=stats::rgamma(n=1,shape=p*a,rate=1/2)
  Ti=rep(1,p)
  beta=rep(0,p)
  hi=rep(1,p)
  ti=1
  betamatrix<-matrix(rep(NA,nmc*p),nrow=5000)
  likelihoodout <- matrix(0, n, nrow=5000)

  #Niter iterations
```

```
for(i in 1:(burn+nmc)){
  #step1:sample sigma^2
  s=c(psi*phi^2*tau^2)
  E_1=max(t(y-x%*%beta)%*%(y-x%*%beta),1e-8)
  E_2=max(sum(beta^2*s),1e-8)
  sigma2=1/stats::rgamma(1,(n+p)/2,rate=(E_1+E_2)/2)
  sigma1=sqrt(sigma2)
  if(sigma1>1e20) print("Please choose a better hyperparameter, it is too big")

  #step2:sample beta
  u=stats::rnorm(p)*sqrt(s)
  delta=stats::rnorm(n)
  v=x%*%u+delta
  stx=as.numeric(s)*t(x)
  w=ginv(x%*%stx+diag(n))%*%(y/sigma1-v)
  beta=(u+(stx%*%w))*sigma1

  mix1=abs(beta)/sigma1
  mix2=mix1/c(phi)
  #step3:sample psi
  mu=tau/mix2
  pv=(stats::rnorm(p))^2
  pu=stats::runif(p)
  temp2=mu*pv
  temp3=sqrt(4*temp2+temp2^2)
  temp4=mu+0.5*(pv*(mu^2))-0.5*(mu*temp3)
  locs=(pu<=mu/(mu+temp4))
  psi1=locs*temp4+(1-locs)*(mu^2/temp4)
  psi=1/psi1
```

```
    #step4:sample tau
    tau=GIGrvg::rgig(n=1,lambda=p*a-p,psi=1,chi=2*sum(mix2))

    #step5:sample phi
    hu=stats::runif(p,0,exp(-1/(2*hi)))
    hl=1/(2*log(1/hu))
    hf=stats::pgamma(hl,shape=1-a,rate=mix1)
    hr=pmin(runif(p,hf,1),(1-(1e-20)))
    hi=stats::qgamma(hr,shape=1-a,rate=mix1)
    Ti=1/hi
    phi=Ti/sum(Ti)
    phi[phi<=(1e-20)]=(1e-20)

    #likelihood
    likelihood <- stats::dnorm(y, mean = x %*% beta, sd = sqrt(sigma2))

    #beta output
    if(i>burn&&i%%thin==0) betamatrix[(i-burn)/thin,]=beta
    if(i%%1000==0) print(i)

    #likelihood output
    if(i>burn&&i%%thin==0) likelihoodout[(i - burn)/thin,] = likelihood
    if(i%%1000==0) print(i)

  }
  result = list( "betaSamples" = betamatrix, "LikelihoodSamples" = likelihoodout)
  return(result)
}
```

Bayesian variable selection via penalized credible region

```
dlvs<-function(dlresult){
  #use penalized credible region to do variable selection
  #matrix computation to make the solutions be accomplished by LASSO
  p=ncol(dlresult)
  betamean<-apply(dlresult,2,mean)
  betacov<-cov(dlresult)
  D<-betamean^2
  cov1=expm::sqrtm(betacov)
  covinv<-MASS::ginv(cov1)
  #scale
  xstar=t(as.numeric(D)*covinv)
  ystar=covinv%*%betamean
  xstar<-scale(xstar, scale=F)
  ystar<-ystar-mean(ystar)
  #solve LASSO problem by glmnet package
  model<-glmnet::glmnet(xstar,ystar,standardize=FALSE,alpha=1,family="gaussian")
  lam=glmnet::cv.glmnet(xstar,ystar,standardize=FALSE,alpha=1,family="gaussian")$lamb
  betastar=coef(model,s=lam)[2:(p+1)]
  betatil=D*betastar
  result1 = list("xstar"=xstar,"ystar"=ystar,"betatil"=betatil)

  return(result1)
}
```

MCMC sampling result by computing the posterior mean, median and credible intervals

```
dlanalysis<-function(dlresult,alpha=0.05){
```

```
  betamean = apply(dlresult,2,mean)
  betamedian = apply(dlresult,2,median)
  min = apply(dlresult,2,quantile,prob=alpha/2)
  max = apply(dlresult,2,quantile,prob=1-alpha/2)
  result = list("betamean"=betamean,"betamedian"=betamedian,"LeftCI"= min,"RightCI"=ma
  return(result)
}
```

Simulation Scheme code

```
y.sd   <- 1     # standard deviation of the data

p      <- 1000    # number of covariates

n      <- 100    # number of samples

beta   <- as.vector(smoothmest::rdoublex(p))  # from double exponential distribution

x <- mvtnorm::rmvnorm(n, mean = rep(0, p))  # from multivariate normal distribution
y.mu   <- x %*% beta  # mean of the data

y <- as.numeric(stats::rnorm(n, mean = y.mu, sd = y.sd))  # from normal distribution

T <- exp(y)   # AFT model

C <- rgamma(n, shape = 1.75, scale = 3)  # censoring time

time <- pmin(T, C)  # observed time is min of censored and true

status = time == T   # set to 1 if event is observed
```

```
ct <- as.matrix(cbind(time = time, status = status))  # censored time
```

ProQuest Number: 29996908

INFORMATION TO ALL USERS
The quality and completeness of this reproduction is dependent on the quality
and completeness of the copy made available to ProQuest.

Distributed by ProQuest LLC (2022).
Copyright of the Dissertation is held by the Author unless otherwise noted.

This work may be used in accordance with the terms of the Creative Commons license
or other rights statement, as indicated in the copyright statement or in the metadata
associated with this work. Unless otherwise specified in the copyright statement
or the metadata, all rights are reserved by the copyright holder.

This work is protected against unauthorized copying under Title 17,
United States Code and other applicable copyright laws.

Microform Edition where available © ProQuest LLC. No reproduction or digitization
of the Microform Edition is authorized without permission of ProQuest LLC.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346 USA