

# Nonparametric Bayesian Lomax delegate racing for survival analysis with competing risks

Quan Zhang and Mingyuan Zhou

Mcombs School of Business, University of Texas at Austin

## Abstract

We propose Lomax delegate racing (LDR) to explicitly model the mechanism of survival under competing risks and to interpret how the covariates accelerate or decelerate the time to event. LDR explains non-monotonic covariate effects by racing a potentially infinite number of sub-risks, and consequently relaxes the ubiquitous proportional-hazards assumption which may be too restrictive. Moreover, LDR is naturally able to model not only censoring, but also missing event times or event types. For inference, we develop a Gibbs sampler under data augmentation for moderately sized data, along with a stochastic gradient descent maximum a posteriori inference algorithm for big data applications. Illustrative experiments are provided on both synthetic and real datasets, and comparison with various benchmark algorithms for survival analysis with competing risks demonstrates distinguished performance of LDR.

## Exponential racing and application to survival analysis

**Property 1** (Exponential racing). *If  $t_j \sim \text{Exp}(\lambda_j)$ ,  $j = 1, \dots, J$ , are independent to each other, then the argument of the minimum  $y = \text{argmin}_j t_j$  and  $t = \min\{t_1, \dots, t_J\}$  satisfy*

$$y \sim \text{Categorical}\left(\lambda_1 / \sum_{j=1}^J \lambda_j, \dots, \lambda_J / \sum_{j=1}^J \lambda_j\right) \\ t \sim \text{Exp}\left(\sum_{j=1}^J \lambda_j\right), \text{ and } y \perp t.$$

- A race among  $J$  teams. Team  $j$ 's completion time  $t_j \sim \text{Exp}(\lambda_j)$ .
- The winner is the team with the minimum completion time.
- Property 1 shows the winner's completion time  $t$  still follows an exponential distribution and is independent of which team wins the race.
- Fully factorized likelihood  $P(y, t | \{\lambda_j\}_{1,J}) = \lambda_y e^{-t \sum_{j=1}^J \lambda_j}$ .

## Survival analysis with competing risks

Exponential racing describes a natural mechanism of surviving under competing risks.

- Team  $j \rightarrow$  competing risk  $j$ ,
- Completion time  $t_j \rightarrow$  the latent survival time under risk  $j$ .
- Winner's completion time  $t \rightarrow$  the observed time to event (or failure time).
- The winning team  $y \rightarrow$  the event type (or cause of failure).
- Easy to deal with censoring.  
If we don't observe  $t$  but know  $t \in \Psi \subseteq \mathbb{R}_+$  with  $P(t \in \Psi | \{\lambda_j\}_{1,J}) = \int_{\Psi} \lambda e^{-\lambda u} du$ , then we draw  $t \sim \text{Exp}_{\Psi}(\sum_j \lambda_j)$  with PDF  $f_{\Psi}(t | \lambda) = \lambda e^{-\lambda t} \delta_{t \in \Psi} / \int_{\Psi} \lambda e^{-\lambda u} du$ , resulting in the likelihood

$$P\left(t, t \in \Psi \mid \sum_j \lambda_j\right) = f_{\Psi}\left(t \mid \sum_j \lambda_j\right) P\left(t \in \Psi \mid \sum_j \lambda_j\right) \\ = \left(\sum_j \lambda_j\right) e^{-t \sum_j \lambda_j}.$$

## Covariate-dependent Lomax distribution

If  $t \sim \text{Exp}(\lambda)$ ,  $\lambda \sim \text{Gamma}(r, 1/b)$  with  $\mathbb{E}(\lambda) = r/b$ , then  $t | r, b \sim \text{Lomax}(r, b)$  with PDF  $f(t | r, b) = \frac{rb^r}{(t+b)^{r+1}}$ ,  $t \in \mathbb{R}_+$ .

We can model covariate dependence of  $t$  by assuming

$$t | r, \mathbf{x}, \boldsymbol{\beta} \sim \text{Lomax}(r, e^{-\mathbf{x}'\boldsymbol{\beta}}),$$

where  $\boldsymbol{\beta}$  and  $\mathbf{x}$  are the regression-coefficient and covariate vectors, respectively. The ease to infer  $\boldsymbol{\beta}$  becomes clear if we express the likelihood as

$$\text{Lomax}(t; r, e^{-\mathbf{x}'\boldsymbol{\beta}}) = \frac{r e^{-r\mathbf{x}'\boldsymbol{\beta}}}{(t + e^{-\mathbf{x}'\boldsymbol{\beta}})^{r+1}} = t^{-1} \text{NB}\left(1; r, \frac{e^{\mathbf{x}'\boldsymbol{\beta} + \ln t}}{1 + e^{\mathbf{x}'\boldsymbol{\beta} + \ln t}}\right)$$

## Lomax delegate racing (LDR)

**Definition 1** (Lomax delegate racing). *Given a random draw of a gamma process  $G_j \sim \text{GP}(G_{0j}, 1/c_{0j})$ , expressed as  $G_j = \sum_{k=1}^{\infty} r_{jk} \delta_{\beta_{jk}}$ , for each  $j \in \{1, \dots, J\}$ , Lomax delegate racing (LDR) models  $y$ , the team number of the winner, and  $t$ , the winner's completion time, given the covariates  $\mathbf{x}$  as*

$$y = \text{argmin}_{j \in \{1, \dots, J\}} t_j, \quad t = \min_{j \in \{1, \dots, J\}} t_j, \quad t_j = t_{j\kappa_j}, \\ \kappa_j = \text{argmin}_{k \in \{1, \dots, \infty\}} t_{jk}, \quad t_{jk} \sim \text{Lomax}(r_{jk}, e^{-\mathbf{x}'\boldsymbol{\beta}_{jk}}).$$

LDR can be interpreted as a two-phase race between multiple teams consisting of countably infinite players.

- In the first phase, the players within a team compete and the completion time of the team is defined as that of its fastest player.
- In the second phase, the teams compete and the winner of the race is the team that has the shortest completion time.  
Survival: teams  $\rightarrow$  competing risks; players  $\rightarrow$  sub-risks.  
 $t$  and (or)  $y$  are (is) observed but  $t_j$ 's are not.

## Inference

- Gibbs sampler  
Introduce an auxiliary variable  $t_i \sim \text{Exp}\left(\sum_{j=1}^J \sum_{k=1}^{\infty} \lambda_{ijk}\right)$ , or  $y_i \sim \text{Categorical}(\lambda_1 / \sum_{j=1}^J \lambda_j, \dots, \lambda_J / \sum_{j=1}^J \lambda_j)$  for individual  $i$ . Gibbs sampler updates are similar as those for negative binomial regression, since
$$P(t_i, \kappa_{iy_i}, y_i | \mathbf{x}_i, \{\boldsymbol{\beta}_{jk}\}_{jk}) \\ = t_i^{-1} \prod_j \prod_k \text{NB}\left(\mathbf{1}(\kappa_{iy_i} = k, y_i = j); r_{jk}, \frac{e^{\mathbf{x}'_i \boldsymbol{\beta}_{jk} + \ln t_i}}{1 + e^{\mathbf{x}'_i \boldsymbol{\beta}_{jk} + \ln t_i}}\right).$$
- Maximum a posteriori (MAP) estimations  
reparameterization:  $\lambda_{ijk} \sim \text{Gamma}(r_{ijk}, e^{\mathbf{x}'_i \boldsymbol{\beta}_{jk}})$   
 $\rightarrow \lambda_{ijk} = \tilde{\lambda}_{ijk} e^{\mathbf{x}'_i \boldsymbol{\beta}_{jk}}$  where  $\tilde{\lambda}_{ijk} \sim \text{Gamma}(r_{ijk}, 1)$ .  
Score function gradients:  $\nabla_{\mathbf{r}} p(\tilde{\boldsymbol{\lambda}}_i | \mathbf{r}) = p(\tilde{\boldsymbol{\lambda}}_i | \mathbf{r}) \nabla_{\mathbf{r}} \log p(\tilde{\boldsymbol{\lambda}}_i | \mathbf{r})$

$$\nabla_{\mathbf{r}} \log p_i = \frac{\int \nabla_{\mathbf{r}} \left[ (p_{t_i} \times p_{y_i}) p(\tilde{\boldsymbol{\lambda}}_i | \mathbf{r}) \right] d\tilde{\boldsymbol{\lambda}}_i}{\int (p_{t_i} \times p_{y_i}) p(\tilde{\boldsymbol{\lambda}}_i | \mathbf{r}) d\tilde{\boldsymbol{\lambda}}_i} \\ = \frac{\int (p_{t_i} \times p_{y_i}) \nabla_{\mathbf{r}} \log p(\tilde{\boldsymbol{\lambda}}_i | \mathbf{r}) p(\tilde{\boldsymbol{\lambda}}_i | \mathbf{r}) d\tilde{\boldsymbol{\lambda}}_i}{\int (p_{t_i} \times p_{y_i}) p(\tilde{\boldsymbol{\lambda}}_i | \mathbf{r}) d\tilde{\boldsymbol{\lambda}}_i}$$

## Contact Information:

Department of Information, Risk and Operations Management

2110 Speedway, Austin, TX 78705, US

Email: quan.zhang@mcombs.utexas.edu



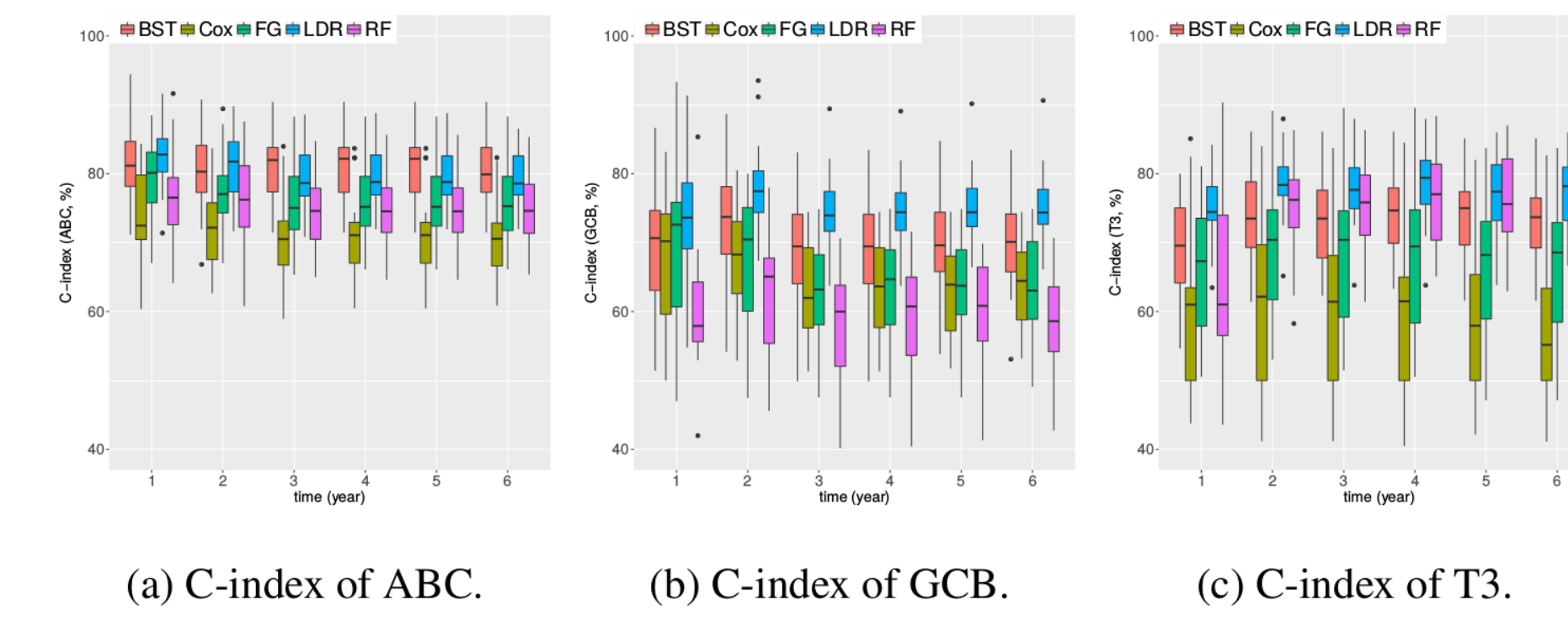
## Experimental results

Table 1: Synthetic data generating process.

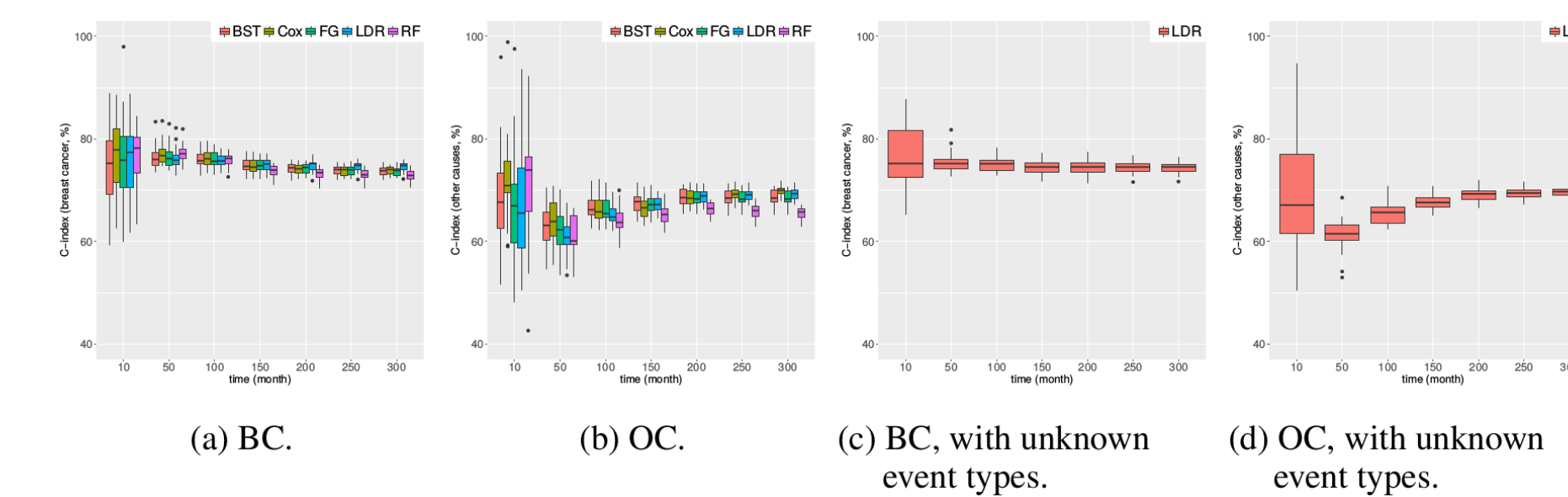
Synthetic data 1		Synthetic data 2	
$t_i = \min(t_{i1}, t_{i2}, 3.5)$ , $t_{i1} \sim \text{Exp}(e^{\mathbf{x}'_i \boldsymbol{\beta}_1})$ , $t_{i2} \sim \text{Exp}(e^{\mathbf{x}'_i \boldsymbol{\beta}_2})$		$t_i = \min(t_{i1}, t_{i2}, 6.5)$ , $t_{i1} \sim \text{Exp}(1/\cosh(\mathbf{x}'_i \boldsymbol{\beta}_1))$ , $t_{i2} \sim \text{Exp}(1/ \sinh(\mathbf{x}'_i \boldsymbol{\beta}_2) )$	

(a) C-index of risk 1 for synthetic data 1. (b)  $r_{jk}$  by descending order for synthetic data 1. (c) C-index of risk 1 for synthetic data 2. (d)  $r_{jk}$  by descending order for synthetic data 2.

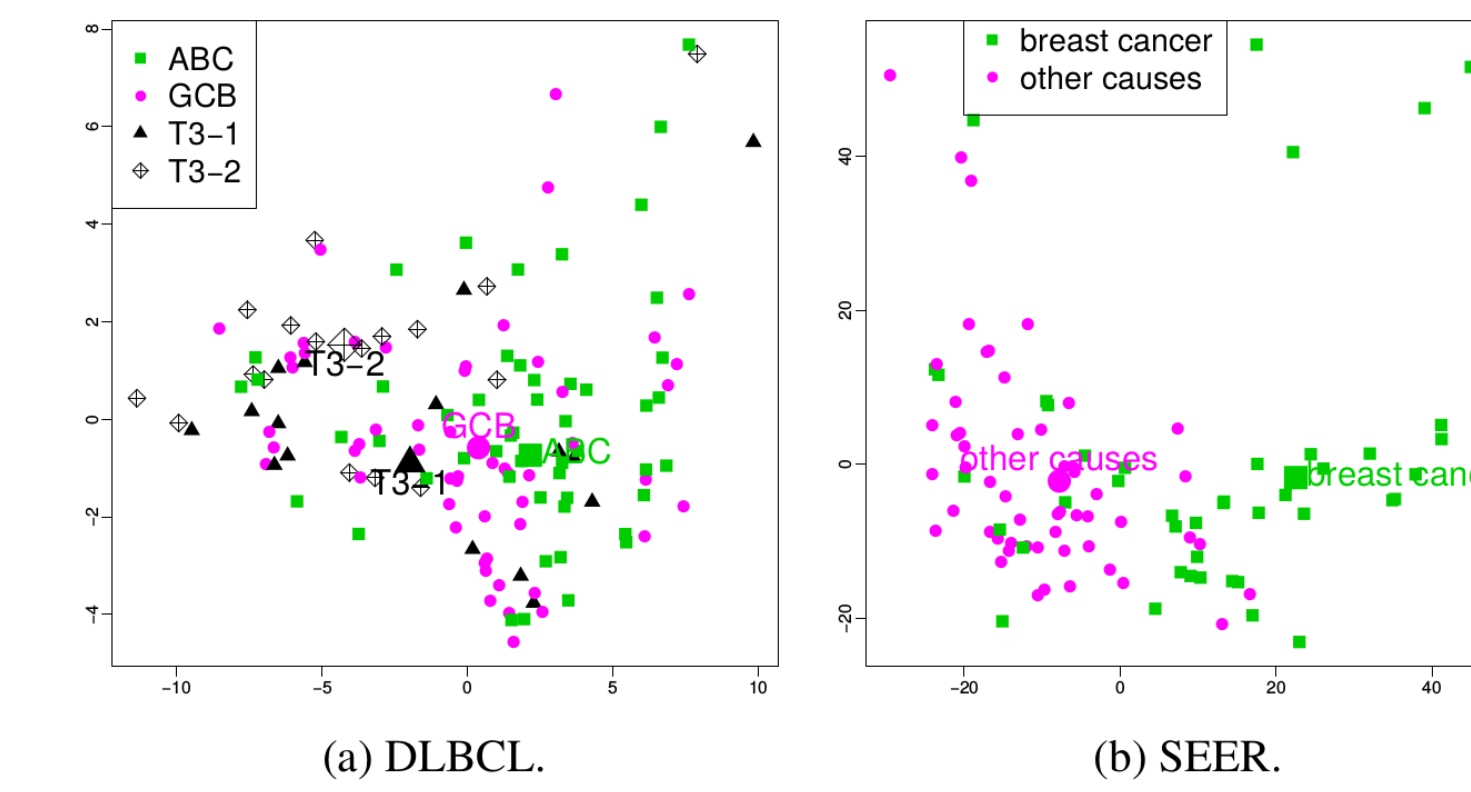
**Figure 1:** Cause-specific C-indices for synthetic data and shrinkage of  $r_{jk}$  by LDR.



**Figure 2:** Cause-specific C-indices for the DLBCL data.



**Figure 3:** Cause-specific C-indices for the SEER breast cancer data.

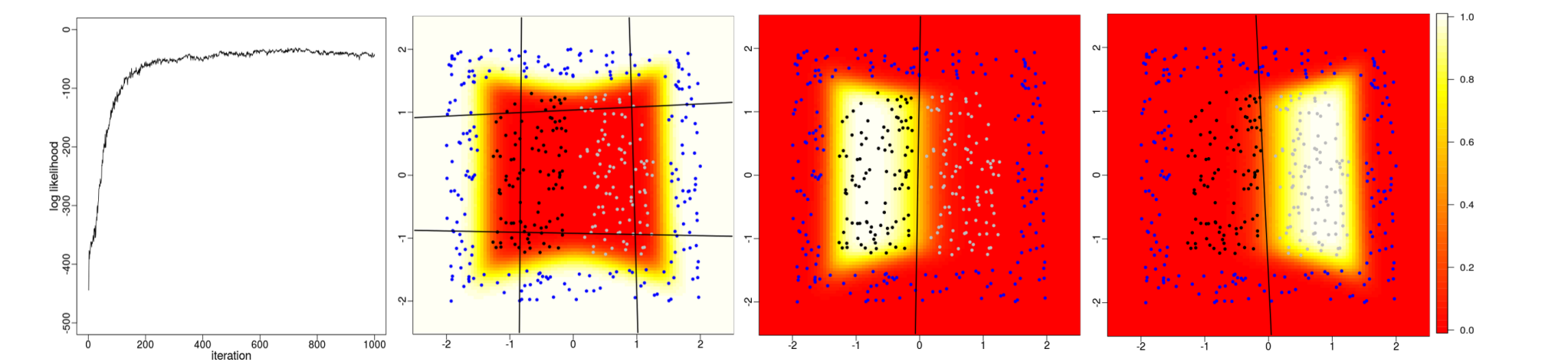


**Figure 4:** Isomap visualization of the observations and inferred sub-risk representations.

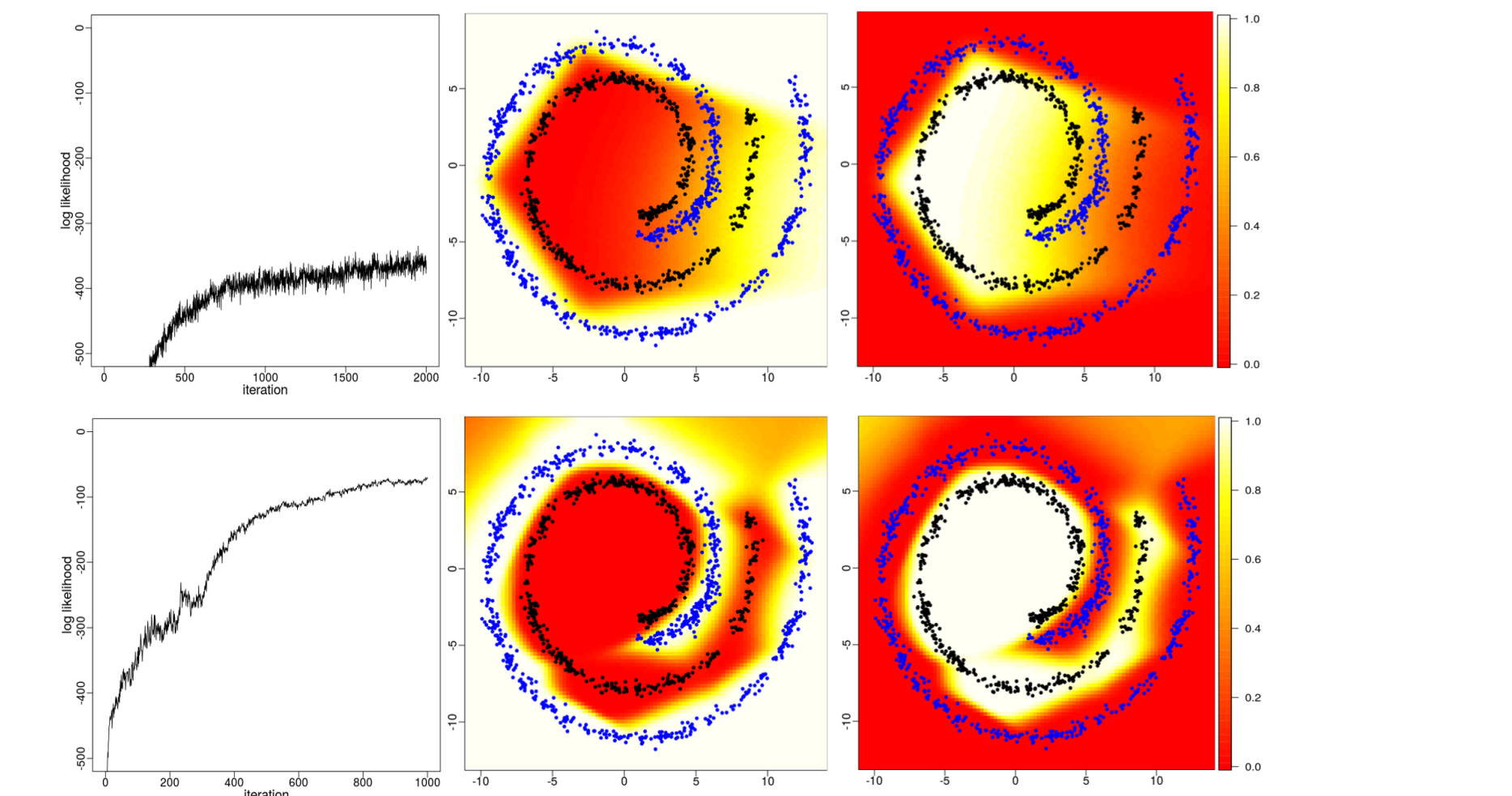
Table 1: Brier score for T3 of DLBCL.

	$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$	$\tau = 5$	$\tau = 6$
Cox	.193±.053	.190±.061	.206±.069	.220±.071	.233±.068	.245±.072
FG	.183±.051	.186±.062	.195±.067	.212±.069	.230±.070	.234±.069
BST	.169±.046	.172±.044	.177±.049	.185±.046	.185±.047	.193±.048
RF	.117±.045	.151±.046	.157±.043	.169±.049	.180±.051	.185±.052
LDR	.111±.035	.137±.038	.142±.036	.151±.041	.165±.044	.171±.046

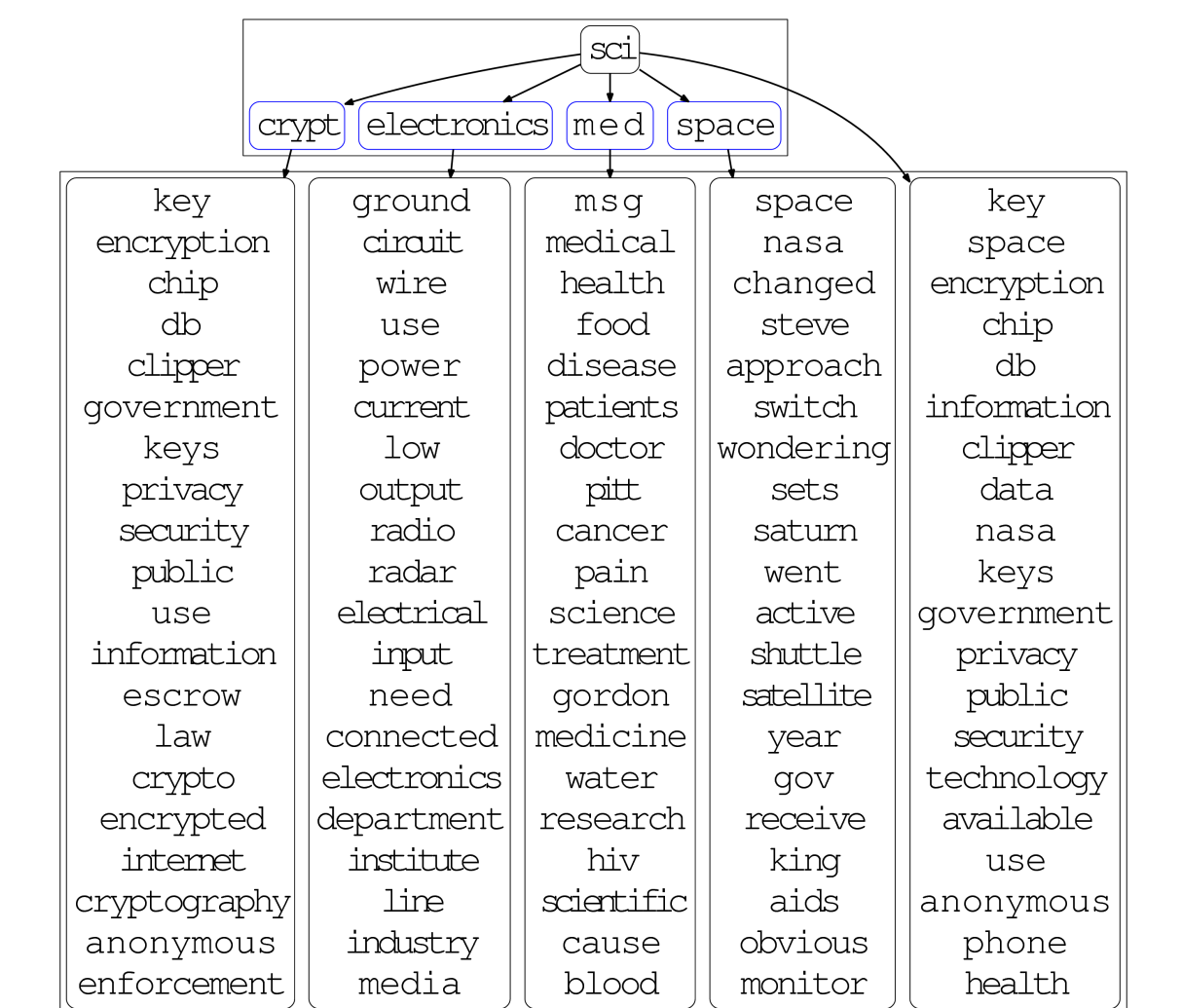
## Additional results for (hierarchical) classification



**Figure 5:** Trajectory plots of log likelihood (column 1) and predictive probability heat maps and the unpruned hyperplanes (column 2 to 4) for *square*.



**Figure 6:** Trajectory plot of log likelihood and predictive probability heat map for *swiss roll* data. Row 1 uses the original data and row 2 the softplus transformation  $\tilde{\mathbf{x}}_i := \left[ \mathbf{x}'_i, \log(1 + e^{\mathbf{x}'_i \boldsymbol{\beta}_{11}}), \dots, \log(1 + e^{\mathbf{x}'_i \boldsymbol{\beta}_{jk}}), \dots, \log(1 + e^{\mathbf{x}'_i \boldsymbol{\beta}_{SK}}) \right]'$ .



**Figure 7:** Top 20 words in subtypes of each node of science in 20 news groups.

## Conclusions

- Introducing covariate dependencies to exponential racing.
- Interpretable nonlinearity for survival analysis and classification.
- Data mining of sub-risks.
- Gibbs sampler by data augmentation.
- Scalability by reparameterization and score function gradients for MAP estimations.