C-index targeted Bayesian methods for personalized risk prediction and uncertainty assessment

1 Things to Do:

For August 10:

- Try running posterior sampling of $G(Y_i)$ again with $\kappa = 0.1$ and delta_alpha $= 10$.
- Double check Cox simulation code. Try with $n = 5,000$.
- Run all simulation results with a standardized design matrix.
- Try adding glmnet results to Cox simulation.
- Try running the hyperparameter learning code on the updated simulations.

Remaining things for project:

- Experiment more with hyperparameter tuning method.
- Construct 1 more simulation scenarios: Use CoxPH and AFT model to start, with a more complex "regression function".
- Measures of performance for simulations: (1) C-statistic for the posterior mean of $\boldsymbol{\theta}$, (2) rank concordance between $\boldsymbol{\theta}$ and rankings of "true" risk scores, (3) Confidence interval for the credible interval for the c-statistic vs. the "true" c-statistic, (4) confidence intervals for ranks of risk scores and their coverage, (5) performance on other measures,
- How to combine this approach within a larger model? (for example, where we have other types of outcomes in addition to survival)
- Real data example to use: (1) breast cancer, (2) SEER data.

2 Notation

We let T_i denote the time to the outcome of interest and let C_i denote the time to censoring for the i^{th} individuals. Then, $Y_i = \min\{T_i, C_i\}$ and $\delta_i = \mathbf{1}\{T_i \leq C_i\}$ will denote the will denote the follow-up time and event indicator for the i^{th} individual respectively. In addition, we observe a p-dimensional covariate vector $\mathbf{x}_i \in \mathbb{R}^p$ for each i. The entire collection of observations will be denoted by $\mathcal{D} = \{(Y_i, \delta_i, \mathbf{x}_i); i = 1, \ldots, n\}$. We let $S(t \mid \mathbf{x}_i)$ and $G(t \mid \mathbf{x}_i)$ denote the survival functions for T_i and C_i conditional on $\mathbf{x}_i$ respectively; that is, $S(t|\mathbf{x}_i) = P(T_i > t \mid \mathbf{x}_i)$ and $G(t|\mathbf{x}_i) = P(C_i > t \mid \mathbf{x}_i)$. We assume that the failure times T_i and censoring times C_i are conditionally independent given $\mathbf{x}_i$.

Our goal is to report inferences for a collection of risk score parameters $\theta_1, \ldots, \theta_n$ with θ_i denoting the risk score parameter for patient i. We will assume that the risk scores $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ can be expressed in terms of another p-component parameter vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ through an $n \times p$ matrix $\mathbf{A}$, i.e., where it is assumed that $p \leq n$. More specifically, we assume that $\boldsymbol{\theta}$ is a normalized version of $\mathbf{A}\boldsymbol{\beta}$:

$$\boldsymbol{\theta} = \frac{\mathbf{A}\boldsymbol{\beta} - \mathbf{1}^T\mathbf{A}\boldsymbol{\beta}/n}{(\mathbf{A}\boldsymbol{\beta} - \mathbf{1}^T\mathbf{A}\boldsymbol{\beta}/n)^T(\mathbf{A}\boldsymbol{\beta} - \mathbf{1}^T\mathbf{A}\boldsymbol{\beta}/n)} \tag{1}$$

[[Assume that the prior for $\boldsymbol{\beta}$ always has mean zero and has average variance 1.]]

Inference for $\boldsymbol{\theta}$ will be driven by a C-index type loss function whose value only depends on the signs of the pairwise differences $\theta_i - \theta_j$. Each loss function $C_{\tau_G}(\boldsymbol{\theta}, G; \mathcal{D})$ we consider is assumed to have the following form

$$\begin{aligned} C_{\tau_G}(\boldsymbol{\theta}, G; \mathcal{D}) &= \sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}(G, \tau_G)\mathbf{1}\{\theta_i > \theta_j\} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}(G, \tau_G)\mathbf{1}\{\mathbf{d}_{i,j}^T\boldsymbol{\theta} > 0\}, \end{aligned} \tag{2}$$

where in (2), $\mathbf{d}_{i,j}$ is the $n \times 1$ "differencing" vector whose elements are constructed so that $\mathbf{d}_{i,j}^T\boldsymbol{\theta} = \theta_i - \theta_j$ for every pair (i, j). The terms $w_{i,j}(G, \tau_G) \geq 0$ in (2) are weights which can depend on the observed patient outcomes, the censoring distribution G, and a truncation

point $\tau_G > 0$, but the weights do not depend on the vector of risk scores parameters $\boldsymbol{\theta}$. The truncation point τ_G is often assumed to be pre-specified and to not depend on G, but we have used included the G subscript in τ_G to allow the truncation point to explicitly depend on certain features of the censoring distribution. For example, using an estimate of a truncation point τ_G which represents the upper end of the support of the censoring distribution was suggested by Tian et al. (2020) in the context of selecting the window for the restricted mean survival time.

A "smooth" version of C_τ in (2) is the following:

$$\tilde{C}_{\tau_G,\nu}(\boldsymbol{\theta}, G; \mathcal{D}) = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{w_{ij}(G, \tau_G)}{1 + \exp(-\mathbf{d}_{i,j}^T \boldsymbol{\theta}/\nu)},$$

2.1 Choices for the C index

If one has a risk score $f(\mathbf{x}_i)$ which takes a patient covariate as input and outputs some measure of risk, a C-index, in the context of survival analysis, can typically be thought of as an estimate of the following "concordance" parameter:

$$\theta_C(f, \tau) = P\big\{f(\mathbf{x}_i) > f(\mathbf{x}_j) | T_j > T_i, T_i \leq \tau\big\}.$$

Several well-known examples of C-indeces which can be expressed in the form (2) are listed below.

Harrell's C-index. An estimate of $\theta_C(f, \tau)$ proposed in Harrell et al. (1982) and further developed in Pencina and D'Agostino (2004) is the following

$$\hat{C}_{\tau_G} = \frac{\sum_{i \neq j} \delta_i I(Y_i < Y_j, Y_i \leq \tau) I\big(f(\mathbf{x}_i > f(\mathbf{x}_j)\big)}{\sum_{i \neq j} \delta_i I(Y_i < Y_j, Y_i \leq \tau_G)}. \tag{3}$$

[[Note that the original Harrell paper did not consider the truncation point]]. [[The above is just a truncated version of Harrell's]]. [[It seems the truncated version of Harrell's c-statistic was considered more in Pencina and D'Agostino (2004)]].

Harrell's C-index can be expressed in the form (2) with the following choice of weight function

$$w_{i,j}(G, \tau_G) = \frac{\delta_i \mathbf{1}\{Y_i < Y_j, Y_i < \tau_G\}}{\sum_{i=1}^n \sum_{j=1}^n \delta_i \mathbf{1}\{Y_i < Y_j, Y_i < \tau_G\}} \tag{4}$$

Uno's C-index. Uno C-index (Uno et al. (2011)) was proposed as an alternative "estimator" of the concordance parameter $\theta_C(f, \tau)$ in (3). The weighting scheme is just different so that it does not converge to something which depends on the censoring distribution.

$$\hat{C}_\tau = \frac{\sum_{i=1}^n \sum_{j=1}^n \delta_i \{G(Y_i)\}^{-2} \mathbf{1}\{Y_i < Y_j, Y_i < \tau\} \mathbf{1}\{f(\mathbf{x}_i) > f(\mathbf{x}_j)\}}{\sum_{i=1}^n \sum_{j=1}^n \delta_i \{G(Y_i)\}^{-2} \mathbf{1}\{Y_i < Y_j, Y_i < \tau\}} \tag{5}$$

Uno's C-index can be expressed in the form (2) with the following choice of weight function

$$w_{i,j}(G, \tau) = \frac{\delta_i \{G(Y_i)\}^{-2} \mathbf{1}\{Y_i < Y_j, Y_i < \tau\}}{\sum_{i=1}^n \sum_{j=1}^n \delta_i \{G(Y_i)\}^{-2} \mathbf{1}\{Y_i < Y_j, Y_i < \tau\}} \tag{6}$$

Time-Dependent ROC. The true underlying cumulative-dynamic time-dependent area under the ROC curve (AUC) is defined at time t as

$$AUC(t) = P\big\{f(\mathbf{x}_i) > f(\mathbf{x}_j) | T_i \leq t, T_j > t\big\} \tag{7}$$

[The formulas are really in Blanche et al. (2013) and Hung and Chiang (2010)]

This is estimated by

$$\hat{AUC}(t) = \frac{1}{n^2 \hat{S}(t)[1 - \hat{S}(t)]\hat{G}(t)} \sum_{i=1}^n \sum_{j=1}^n \frac{\delta_i}{\hat{G}(Y_i)} I(Y_i \leq t, Y_j > t) I(f(\mathbf{x}_i) > f(\mathbf{x}_j)) \tag{8}$$

For a single score, we will typically evaluate the biomarker by integrating the function

$AUC(t)$ across t

$$\begin{aligned}\widehat{CAUC} &= \int_0^\tau W(t)\hat{AUC}(t)dt \\ &= \int_0^\tau \frac{W(t)}{n^2\hat{S}(t)[1-\hat{S}(t)]\hat{G}(t)} \sum_{i=1}^n \sum_{j=1}^n \frac{\delta_i}{\hat{G}(Y_i)} I(Y_i \le t, Y_j > t) I(f(\mathbf{x}_i) > f(\mathbf{x}_j))dt\end{aligned}$$

Use the fact that

$$\hat{S}(t) = \frac{1}{n}\sum_{i=1}^n \frac{I(Y_i \le t)\delta_i}{\hat{G}(Y_i)}$$

Here, we have

$$\begin{aligned}w_{i,j}(G,\tau) &\propto \frac{\delta_i}{\hat{G}(Y_i)} \int_0^\tau \frac{W(t)}{n^2\hat{S}(t)[1-\hat{S}(t)]\hat{G}(t)} I(Y_i \le t, Y_j > t)dt \\ &= \frac{\delta_i}{n^2\hat{G}(Y_i)} \sum_{k=1}^K \frac{I(Y_i \le t_k, Y_j > t_k)}{\hat{S}(t_{k-1})[1-\hat{S}(t_k)]\hat{G}(t_{k-1})} \int_{t_{k-1}}^{t_k} W(t)dt \end{aligned} \tag{9}$$

$0 = t_0 < t_1 < t_2 < ... < t_{K-1} < t_K = \tau$. To start with, assume $W(t) = 1$ so that $\int_{t_{k-1}}^{t_k} W(t)dt = t_k - t_{k-1}$.

Blanche et al. (2013), Heagerty et al. (2000), Blanche et al. (2019), Uno et al. (2007)

Competing Risks. Wolbers et al. (2014) [[compare with other references]]

Time-Varying Covariates. Is there something that estimates the following landmark-type quantity?

$$U(t_k) = P\Big(f(\mathbf{x}_i(t_k)) > f(\mathbf{x}_j(t_k))\Big| T_i \le t_{k+1}, T_j > t_{k+1}, T_i > t_k, T_j > T_k\Big) \tag{10}$$

$$C = \sum_{k=1}^K U(t_k) \tag{11}$$

What about Gonen's measure?

3 C-index type Loss function and Posterior of Interest

Use framework of Bissiri et al. (2016).

Consider the "soft" version of the C-index:

$$\tilde{C}_{\tau,\nu}(\boldsymbol{\theta}, G; \mathcal{D}) = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{w_{ij}(G,\tau)}{1+\exp(-\theta_{ij}/\nu)},$$

where ν is a relatively small number. In practice, we will use $\nu = 0.22$.

Note that the magnitudes of the risk scores can play a modest role in the soft C index whereas only the rankings of the risk scores would matter for traditional C-indeces.

We will target the loss function which is the negative log of the soft C-statistic (plus a penalty for (remove this)

$$H_{\tau,\nu}(\boldsymbol{\theta}, G; \mathcal{D}) = -\log\{\tilde{C}_{\tau,\nu}(\boldsymbol{\theta}, G; \mathcal{D})\}$$

[[might make sense to consider another transformation of $H_{\tau,\nu}(\boldsymbol{\theta}, G; \mathcal{D})$?]]

In the generalized Bayes framework, our goal is to sample from the "Gibbs posterior"

$$p(\boldsymbol{\beta}|\mathcal{D}, G, \boldsymbol{\lambda}) = \frac{\exp\{-\eta H_{\tau,\nu}(\boldsymbol{\theta}, G, \mathcal{D})\}\pi(\boldsymbol{\beta}|\boldsymbol{\lambda})}{\int\int \exp\{-\eta H_{\tau,\nu}(\boldsymbol{\theta}, G, \mathcal{D})\}\pi(\boldsymbol{\beta}|\boldsymbol{\lambda})d\boldsymbol{\beta}}$$

where $\pi(\boldsymbol{\beta}|\boldsymbol{\lambda})$ is the prior distribution of $\boldsymbol{\beta}$ conditional on the values of additional hyper-parameters $\boldsymbol{\lambda}$.

The main posteriors will be $p(\boldsymbol{\theta}|\mathcal{D})$ which can be directly derived from the posterior $p(\boldsymbol{\beta}|\mathcal{D})$. The posterior $p(\boldsymbol{\beta}|\mathcal{D})$ can be decomposed as

$$p(\boldsymbol{\beta}|\mathcal{D}) = \int p(\boldsymbol{\beta}|G, \mathcal{D})p(G|\mathcal{D})dG$$

3.1 Identifiability Concerns/Interpretation

As we mentioned before, we are assuming that $\boldsymbol{\theta}$ is a normalized version of $\mathbf{A}\boldsymbol{\beta}$.

4 Modeling the Risk scores and prior distributions

Regression Model Examples

Linear Regression. In this class of models, we assume that the vector of risk scores are given by $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta}$ (i.e., $\mathbf{A} = \mathbf{X}$) where $\mathbf{X}$ is an $n \times p$ design matrix that does not contain an intercept column. We place the following prior distribution on $\boldsymbol{\beta}$

$$\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0).$$

Gaussian Processes. To model the risk scores with Gaussian processes, we take $\boldsymbol{\theta} = \boldsymbol{\beta}$, (i.e., $\mathbf{A} = \mathbf{I}_n$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix). The prior distribution of $\boldsymbol{\beta}$ conditional on a vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)$ of hyperparameters is given by

$$\boldsymbol{\beta}|\boldsymbol{\lambda} \sim \text{Normal}(\boldsymbol{\mu}_{\boldsymbol{\lambda}}, \mathbf{K}_{\boldsymbol{\lambda}}), \tag{12}$$

where the (i, j) element of the $p \times p$ matrix $\mathbf{K}_{\boldsymbol{\lambda}}$ is determined by the value of the kernel function for inputs $\mathbf{x}_i$ and $\mathbf{x}_j$, i.e. $K_{\boldsymbol{\lambda}}(\mathbf{x}_i, \mathbf{x}_j)$. A common choice for the kernel function is the squared exponential kernel

$$\mathbf{K}_{\boldsymbol{\lambda}}(\mathbf{x}_i, \mathbf{x}_j) = \exp\Big\{ -\frac{1}{2} \sum_{k=1}^{p} \frac{(x_{ik} - x_{jk})^2}{\lambda_k} \Big\}.$$

In our implementation of the Gaussian process models, we assume that $\boldsymbol{\mu}_{\boldsymbol{\lambda}} = 0$ and that the λ_j have independent Gamma prior distributions, that is,

$$\lambda_1, \ldots, \lambda_p \sim \text{Inverse-Gamma}(\alpha_0, \nu_0).$$

Global-Local Shrinkage Priors. This a class of prior distributions which do not generate uniform-type of shrinkage for the regression coefficients allowing the larger signals to not be shrunk as strongly to zero.

For the choice of global-local shrinkage prior, we focus on the horshoe prior (cite) which assumes that $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta}$ with the p regression coefficients having the following prior distribution

$$\beta_j|\lambda_j, \tau \sim N(0, \tau\lambda_j) \qquad \lambda_j \sim C^+(0,1), \tag{13}$$

where $C^+(0,1)$ denotes the "half-Cauchy" distribution.

5 Prior Specification

5.1 Learning rate and hyperparameter choice

5.2 Prior Distribution and learning rate here

Assume that

$$\pi(\boldsymbol{\beta}|\boldsymbol{\lambda}) \propto |\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}|^{-1/2} \exp\Big\{ -\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\lambda}})^T \boldsymbol{\Sigma}_{\boldsymbol{\lambda}}^{-1/2} (\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\lambda}}) \Big\} \tag{14}$$

where $\mathbf{1}^T\boldsymbol{\mu}_{\boldsymbol{\lambda}} = 0$ and $\text{tr}(\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}) = p$.

For the topic of prior elicitation, it might be helpful to think about defining $\mathbf{y} = \mathbf{P}\boldsymbol{\beta}$, where $\boldsymbol{\Sigma}^{-1} = \mathbf{P}\mathbf{D}\mathbf{P}^T$ so that $\mathbf{y}^T\mathbf{y}$.

6 Posterior Computation

General Outline for an Metropolis-Hastings algorithm for fixed value of G:

1. Propose $\tilde{\boldsymbol{\beta}} \sim \text{Normal}(\boldsymbol{\beta}^{(t)}, \boldsymbol{\Sigma}_{prop,\beta})$.

2. Compute
$$\tilde{\boldsymbol{\theta}} = \frac{\mathbf{A}\tilde{\boldsymbol{\beta}} - \mathbf{1}^T\mathbf{A}\tilde{\boldsymbol{\beta}}/n}{(\mathbf{A}\tilde{\boldsymbol{\beta}} - \mathbf{1}^T\mathbf{A}\tilde{\boldsymbol{\beta}}/n)^T(\mathbf{A}\tilde{\boldsymbol{\beta}} - \mathbf{1}^T\mathbf{A}\tilde{\boldsymbol{\beta}}/n)}$$
and compute the MH ratio
$$rMH_\beta = \frac{\{\tilde{C}_\tau(\tilde{\boldsymbol{\theta}}, G; \mathcal{D})\}^\eta \pi(\tilde{\boldsymbol{\beta}}|\boldsymbol{\lambda}^{(t)})}{\{\tilde{C}_\tau(\boldsymbol{\theta}^{(t)}, G; \mathcal{D})\}^\eta \pi(\boldsymbol{\beta}^{(t)}|\boldsymbol{\lambda}^{(t)})}.$$

3. Generate $u \sim \text{Uniform}(0, 1)$. If $u \leq rMH_\beta$, set $\boldsymbol{\beta}^{(t+1)} = \tilde{\boldsymbol{\beta}}$. Otherwise, if $u > rMH_\beta$, set $\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)}$.

4. Propose $\log \tilde{\boldsymbol{\lambda}} \sim \text{Normal}(\log \boldsymbol{\lambda}^{(t)}, \boldsymbol{\Sigma}_{prop,\lambda})$.

5. Compute the MH ratio
$$rMH_\lambda = \frac{\pi(\boldsymbol{\beta}^{(t+1)}|\tilde{\boldsymbol{\lambda}})\pi(\tilde{\boldsymbol{\lambda}})}{\pi(\boldsymbol{\beta}^{(t+1)}|\boldsymbol{\lambda}^{(t)})\pi(\boldsymbol{\lambda}^{(t)})}.$$

6. Generate $u \sim \text{Uniform}(0, 1)$. If $u \leq rMH_\lambda$, set $\boldsymbol{\lambda}^{(t+1)} = \tilde{\boldsymbol{\lambda}}$. Otherwise, if $u > rMH_\lambda$, set $\boldsymbol{\lambda}^{(t+1)} = \boldsymbol{\lambda}^{(t)}$.

Linear Regression. For linear regression, set
$$\boldsymbol{\Sigma}_{prop,\beta} = \kappa(\mathbf{A}^T\mathbf{A})^{-1}.$$
and choose κ to have a good acceptance ratio.

7 Posterior Computation with Adaptive Hyperparameter Learning

For this section, maybe cite: (Liang et al. (2022))

Before running the MCMC, generate a random train/test split (use 80/20). $\mathbf{A}_{tr}$ is the value of $\mathbf{A}$ on the training set, and $\mathbf{A}_{te}$ is the value of $\mathbf{A}$ on the test set. $\boldsymbol{\theta}_{tr}^{(t)}$ is the value of $\boldsymbol{\theta}^{(t)}$ only on the training observations. $\boldsymbol{\theta}_{test}^{(t)}$ is the value of $\boldsymbol{\theta}^{(t)}$ only on the test observations.

Also, we will run 3 parallel MCMC chains. Let $\boldsymbol{\beta}^{(t),k}$ denote the draw for chain k ($k = 1, 2, 3$) in MCMC iteration t. Let $\boldsymbol{\theta}^{(t),k}$ denote the draw for chain k in MCMC iteration t.

The terms a_t and c_t are defined as $a_t = \frac{1}{t}$ and $c_t = \frac{1}{t^{1/3}}$

One iteration of the MCMC procedure:

1. Propose $\tilde{\boldsymbol{\beta}}_k \sim \text{Normal}(\boldsymbol{\beta}^{(t),k}, \boldsymbol{\Sigma}_{prop,\beta})$, for $k = 1, 2, 3$.

2. Compute
$$\tilde{\boldsymbol{\theta}}_{tr,k} = \frac{\mathbf{A}_{tr}\tilde{\boldsymbol{\beta}}_k - \mathbf{1}^T\mathbf{A}_{tr}\tilde{\boldsymbol{\beta}}_k/n}{(\mathbf{A}_{tr}\tilde{\boldsymbol{\beta}}_k - \mathbf{1}^T\mathbf{A}_{tr}\tilde{\boldsymbol{\beta}}_k/n)^T(\mathbf{A}_{tr}\tilde{\boldsymbol{\beta}}_k - \mathbf{1}^T\mathbf{A}_{tr}\tilde{\boldsymbol{\beta}}_k/n)}$$
and compute the MH ratios
$$rMH_{\beta,k} = \frac{\{\tilde{C}_\tau(\tilde{\boldsymbol{\theta}}_{tr,k}, G; \mathcal{D})\}^{\eta^{(t),k}}\pi(\tilde{\boldsymbol{\beta}}_k|\boldsymbol{\lambda}^{(t),k})}{\{\tilde{C}_\tau(\boldsymbol{\theta}_{tr}^{(t),k}, G; \mathcal{D})\}^{\eta^{(t),k}}\pi(\boldsymbol{\beta}^{(t),k}|\boldsymbol{\lambda}^{(t),k})}.$$

3. Generate $u_k \sim \text{Uniform}(0, 1)$ independently for $k = 1, 2, 3$. If $u_k \leq rMH_{\beta,k}$, set $\boldsymbol{\beta}^{(t+1),k} = \tilde{\boldsymbol{\beta}}_k$. Otherwise, if $u_k > rMH_{\beta,k}$, set $\boldsymbol{\beta}^{(t+1),k} = \boldsymbol{\beta}^{(t),k}$.

4. Propose $\log\tilde{\boldsymbol{\lambda}}_k \sim \text{Normal}(\log\boldsymbol{\lambda}^{(t),k}, \boldsymbol{\Sigma}_{prop,\lambda})$, for $k = 1, 2, 3$.

5. Compute the MH ratios
$$rMH_{\lambda,k} = \frac{\pi(\boldsymbol{\beta}^{(t+1),k}|\tilde{\boldsymbol{\lambda}}_k)\pi(\tilde{\boldsymbol{\lambda}}_k|\alpha_0^{(t),k}, \nu_0^{(t),k})}{\pi(\boldsymbol{\beta}^{(t+1),k}|\boldsymbol{\lambda}^{(t),k})\pi(\boldsymbol{\lambda}^{(t),k}|\alpha_0^{(t),k}, \nu_0^{(t),k})}.$$

6. Generate $u_k \sim \text{Uniform}(0, 1)$, independently for $k = 1, 2, 3$. If $u_k \leq rMH_{\lambda,k}$, set $\boldsymbol{\lambda}^{(t+1),k} = \tilde{\boldsymbol{\lambda}}_k$. Otherwise, if $u_k > rMH_{\lambda,k}$, set $\boldsymbol{\lambda}^{(t+1),k} = \boldsymbol{\lambda}^{(t),k}$.

7. Update hyperparameters as
$$\begin{bmatrix}\alpha_0^{(t+1),1}\\ \nu_0^{(t+1),1}\\ \eta^{(t+1),1}\end{bmatrix} = \begin{bmatrix}\alpha_0^{(t),1}\\ \nu_0^{(t),1}\\ \eta^{(t),1}\end{bmatrix} + a_t\left(\frac{\tilde{C}_\tau(\boldsymbol{\theta}_{test}^{(t+1),2}, G; \mathcal{D}) - \tilde{C}_\tau(\boldsymbol{\theta}_{test}^{(t+1),3}, G; \mathcal{D})}{2c_t}\right)$$

8. Set: $\alpha_0^{(t+1),2} = \alpha_0^{(t+1),1} + c_t$, $\alpha_0^{(t+1),3} = \alpha_0^{(t+1),1} - c_t$,
$\nu_0^{(t+1),2} = \nu_0^{(t+1),1} + c_t$, $\nu_0^{(t+1),3} = \nu_0^{(t+1),1} - c_t$,
$\eta^{(t+1),2} = \eta^{(t+1),1} + c_t$, $\eta^{(t+1),3} = \eta^{(t+1),1} - c_t$.

Linear Regression. For linear regression, set

$$\boldsymbol{\Sigma}_{prop,\beta} = \kappa(\mathbf{A}^T\mathbf{A})^{-1}.$$

and choose κ to have a good acceptance ratio.

8 Simulations

8.1 Basic Cox Model

Repeat the Cox-PH and AFT simulations with more complicated regression function. For the Cox-PH model assume that the log hazard function is

$$\log h(t|\mathbf{x}_i) = -\frac{1}{2}\log(t) + \sum_{j=1}^{10} \beta_j x_{ij}$$

where $\beta_1 = -0.2$, $\beta_2 = -0.1$, $\beta_3 = -0.4$, $\beta_4 = 3.5$, $\beta_5 = 0.4$, $\beta_6 = 1.2$, and $\beta_7 = \beta_8 = \beta_9 = \beta_{10} = 0$. Generate $x_{ij} \sim \text{Normal}(0, 1)$ independently for $j \neq 6$ and generate $x_{i6} = \exp(1 + Z_i)$, where $Z_i \sim \text{Normal}(0, 1)$.

Similarly, for the AFT model assumed that

$$\log T_i = +\sum_{j=1}^{10} \beta_j x_{ij} + \varepsilon_i$$

where $\beta_1 = -0.2$, $\beta_2 = -0.1$, $\beta_3 = -0.4$, $\beta_4 = 3.5$, $\beta_5 = 0.4$, $\beta_6 = 1.2$, and $\beta_7 = \beta_8 = \beta_9 = \beta_{10} = 0$. Generate $x_{ij} \sim \text{Normal}(0, 1)$ independently for $j \neq 6$ and generate $x_{i6} = \exp(1 + Z_i)$, where $Z_i \sim \text{Normal}(0, 1)$.

Rank Concordance

From the estimate of β, compute $\hat{\theta}_i = \mathbf{x}_i^T\hat{\boldsymbol{\beta}}$ for each i in test set. The true rankings are the ranks of $\theta_i = \mathbf{x}_i^T\boldsymbol{\beta}$ for i in test set. Then, look at the rank correlation between $\hat{\theta}_i$ and θ_i,

for i in test set.

```
cor(theta.hat, theta, method="spearman")
cor(theta.hat, theta, method="kendall")
```

To simulate data from these models (inverse transform method). Derive the survival function $S(t|\mathbf{x}_i)$ from $h(t|\mathbf{x}_i)$. Then, set $F(t|\mathbf{x}_i) = 1 - S(t|\mathbf{x}_i)$. Then, draw $U_i \sim Unif(0, 1)$. Then, generate the survival time $T_i = F^{-1}(U_i)$.

8.2 AFT model

The log-hazard is either

$$h(t|\mathbf{x}_i) = h_0(t) \exp\{\mathbf{x}_i^T \boldsymbol{\beta}\} \tag{15}$$

or

9 Example Code

```
library(survival)
data(cancer)

HarrellC <- function(Y, delta, tau) {
    n <- length(Y)
    Wmat <- matrix(0, nrow=n, ncol=n)

    for(i in 1:n) {
        if(delta[i] != 0) {
            Wmat[i,] <- delta[i]*(Y[i] < Y)*(Y[i] < tau)
        }
    }
```

```
    Wmat <- Wmat/sum(Wmat)
    return(Wmat)
}

Wtry <- HarrellC(Y=gbsg$rfstime, delta=gbsg$status, tau=2500)


# Try running the MCMC for linear regression with this
# choice of A
A <- model.matrix(rfstime ~ age + meno + size + grade + er + hormon, data=gbsg)
```

10 Inference for Risk Scores

11 Discussion

Hoff (2009), Mayr and Schmid (2014), Lin and Peng (2013), Han (1987) Gönen and Heller (2005), Wood (1994), Hornik and Grün (2014), He et al. (2022), Schmid et al. (2016), Khan and Tamer (2007), Song et al. (2007), **?**, Wolbers et al. (2009)

References

Bissiri, P. G., C. C. Holmes, and S. G. Walker (2016). A general framework for updating belief distributions. *Journal of the Royal Statistical Society. Series B, Statistical methodology 78*(5), 1103.

Blanche, P., J.-F. Dartigues, and H. Jacqmin-Gadda (2013). Review and comparison of ROC curve estimators for a time-dependent outcome with marker-dependent censoring. *Biometrical Journal 55*(5), 687–704.

Blanche, P., M. W. Kattan, and T. A. Gerds (2019). The c-index is not proper for the evaluation of-year predicted risks. *Biostatistics 20*(2), 347–357.

Gönen, M. and G. Heller (2005). Concordance probability and discriminatory power in proportional hazards regression. *Biometrika 92*(4), 965–970.

Han, A. K. (1987). Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator. *Journal of Econometrics 35*(2-3), 303–316.

Harrell, F. E., R. M. Califf, D. B. Pryor, K. L. Lee, and R. A. Rosati (1982). Evaluating the yield of medical tests. *JAMA 247*(18), 2543–2546.

He, B., S. Ma, X. Zhang, and L. Zhu (2022). Rank-based greedy model averaging for high-dimensional survival data. *Journal of the American Statistical Association* (just-accepted), 1–27.

Heagerty, P. J., T. Lumley, and M. S. Pepe (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics 56*(2), 337–344.

Hoff, P. D. (2009). Simulation of the matrix Bingham–von Mises–Fisher distribution, with applications to multivariate and relational data. *Journal of Computational and Graphical Statistics 18*(2), 438–456.

Hornik, K. and B. Grün (2014). movMF: an R package for fitting mixtures of von Mises-Fisher distributions. *Journal of Statistical Software 58*(10), 1–31.

Hung, H. and C.-t. Chiang (2010). Optimal composite markers for time-dependent receiver operating characteristic curves with censored survival data. *Scandinavian journal of statistics 37*(4), 664–679.

Khan, S. and E. Tamer (2007). Partial rank estimation of duration models with general forms of censoring. *Journal of Econometrics 136*(1), 251–280.

Liang, X., S. Livingstone, and J. Griffin (2022). Adaptive random neighbourhood informed Markov chain Monte Carlo for high-dimensional Bayesian variable selection. *Statistics and Computing 32*(5), 84.

Lin, H. and H. Peng (2013). Smoothed rank correlation of the linear transformation regression model. *Computational Statistics & Data Analysis 57*(1), 615–630.

Mayr, A. and M. Schmid (2014). Boosting the concordance index for survival data – a unified framework to derive and evaluate biomarker combinations. *PloS one 9*(1), e84483.

Pencina, M. J. and R. B. D'Agostino (2004). Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Statistics in medicine 23*(13), 2109–2123.

Schmid, M., M. N. Wright, and A. Ziegler (2016). On the use of Harrell's C for clinical risk prediction via random survival forests. *Expert Systems with Applications 63*, 450–459.

Song, X., S. Ma, J. Huang, and X.-H. Zhou (2007). A semiparametric approach for the nonparametric transformation survival model with multiple covariates. *Biostatistics 8*(2), 197–211.

Tian, L., H. Jin, H. Uno, Y. Lu, B. Huang, K. M. Anderson, and L. Wei (2020). On the empirical choice of the time window for restricted mean survival time. *Biometrics 76*(4), 1157–1166.

Uno, H., T. Cai, M. J. Pencina, R. B. D'Agostino, and L.-J. Wei (2011). On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in medicine 30*(10), 1105–1117.

Uno, H., T. Cai, L. Tian, and L.-J. Wei (2007). Evaluating prediction rules for t-year survivors with censored regression models. *Journal of the American Statistical Association 102*(478), 527–537.

Wolbers, M., P. Blanche, M. T. Koller, J. C. Witteman, and T. A. Gerds (2014). Concordance for prognostic models with competing risks. *Biostatistics 15*(3), 526–539.

Wolbers, M., M. T. Koller, J. C. Witteman, and E. W. Steyerberg (2009). Prognostic models with competing risks: methods and application to coronary risk prediction. *Epidemiology*, 555–561.

Wood, A. T. (1994). Simulation of the von Mises Fisher distribution. *Communications in statistics-simulation and computation 23*(1), 157–164.