# A review of survival trees

1. Introduction

* **Cox proportional hazard regression model** and its extension – useful for studying time-to-event data

>< **Force a specific link between the covariate and response**

* **Interactions between covariates** can be incorporated >< must be specified by analyst
* In practice, inference often made after many models have been tried >< still largely unknown
* When the analyst **does not wish to impose a link function** right from the start, more flexible approaches are available
* **Survival trees and forests** – non-parametric alternatives to semi-parametric models
* Great flexibility
* Can automatically **detect certain types of interactions without the need to specify** them beforehand
* A single tree can naturally group subjects according to their survival behavior based on their covariates **-> easily derive prognostic groups**
* **Ensemble methods** -> powerful predictive tools
* **Development of survival trees** – extend existing tree methods to the case of survival data with censoring
* Treat more complex situations (e.g., multivariate and correlated survival data)
* Study the use of ensemble methods with survival trees
* Deal with time-varying covariates and time-to-event variables measured on a discrete scale

1. **Basic tree building method**

* Classification and regression tree (CART)
* **Basic idea:** **partition covariate space recursively** to form groups of subjects similar according to outcome of interest
* Achieved by minimizing a measure of node impurity (e.g., Gini, entropy, sum of squared deviations)
* Basic approach: binary splits using a single covariate
* Starts at the root node with all obs
* Performs an exhaustive search through all potential binary splits with the covariates
* Selects the best one according to a splitting criterion (e.g., impurity measure)
* CART approach: process repeated until a stopping criterion is met (often until a minimum node size attained)
* Produces a large tree that usually overfits the data
* Pruning and selection method is then applied to find an appropriate subtree
* Alternatively, ensemble of trees can be used – avoids the problem of selecting a single tree
* Appropriate node summaries – usually computed at the terminal nodes to interpret tree or obtain predicted vals
* Node average – typically used for continuous outcome
* Node proportions of each value – for categorical outcome
* Most frequent value at a node can be used if a single prediction is needed
* Survival outcome: KM estimate of the survival function in the node

1. Survival data description

* Observed data composed of – time until either the event occurs or subject is censored
* , vector of p covariates
* Basic setup: Assume that the covariate values are available at time 0 for each subject -> only the baseline values of a time-varying covariate are typically used

1. Survival tree building methods
2. Splitting criteria

* Gordon and Olshen (1985): Force each node to be more homogenous as measured by a Wasserstein metric between the survival function obtain from the KM estimator at the node and a survival function that has mass on at most one finite point
* Ciampi et al (1986): Logrank statistic to compare the 2 groups formed by the children nodes. Retained split is the one with the largest significant test statistic value -> A split which assures the best separation of thee median survival times in the 2 groups
* Ciampi et al (1987): Likelihood ratio statistic (LRS) under an assumed model to measure the dissimilarity between 2 children nodes -> relies on the assumptions related to the chosen model (e.g., Cox PH, exponential model)
* Davis and Anderson (1989): exponential model log-likelihood - equivalent to LRS dissimilarity measure under the exponential model
* Ciampi et al (1988); Ciampi, Thiffault and Sagman (1989): logrank and Wilcoxon-Gehan statistic as dissimilarity measures and KS statistic to compare survival curves of 2 nodes
* Segal (1988): between-node separation (dissimilarity measure) approach based on the Taron-Ware class of 2-sample statistics for censored data (e.g., logrank, Wilcoxon-Gehan statistics)
* Leblanc and Crowley (1993): logrank statistic with new method for pruning and selecting a final tree based on measure of split-complexity
* Therneau, Grambsch and Fleming (1990): martingale residuals from a null Cox model used as outcome for a regression tree algorithm
* Keles and Segal (2002): analytic relationship between the logrank and martingale residuals sum of squares split functions
* Loh (1991) and Ahn and Loh (1994): 2 splitting criteria based on residuals obtained from fitting a Cox model with one covariate at a time – study the patterns of the Cox model residuals along each covariate axis and then select the splitting covariate whose axis patterns appear the least random (degree of randomness of residuals = observations in parent node divided into 2 classes along each covariate and measured by 2-sample t-test)
* Leblanc and Crowley (1992): Exploiting and equivalence between the proportional hazard, full likelihood model and a Poisson likelihood model -> splitting criterion based on a node deviance measure between a saturated model log-likelihood and a maximized log-likelihood. Unknown full likelihood is approximated by replacing the baseline cumulative hazard function by the Nelson-Aalen estimator -> easy implementation
* Zhang (1995): impurity criterion which combines 2 separate impurity measures, on for observed times and one for proportion of censored observations
* Molinaro, Dudoit and van der Laan (2004): defining an observed data-world (with censoring) loss function by weighting a full data-world (without censoring) loss function. Each non-censored observation is weighted by the inverse prob of censoring (IPC) given the covariates
* Jin et al (2004): splitting rule basedon the variance of survival times >< mean and variance survival times affected by censored observations -> using a restricted time limit to compute the variance
* Cho and Hong (2008): L1 loss function to build a median survival tree. Censored observation

1. Selection of a single tree:

* Decide when to stop splitting -> select a specific tree as the final model
* Too large -> overfitting
* Too small -> underfitting
* 2 approach to select final tree:
* Backward method: Large tree -> pruning
* Forward method: built-in stopping rule to decide when to stop splitting a node further
* (more in Appendix)

1. Some variants and related methods (See paper for more info)
2. Comparison of methods (See paper for more info)
3. Ensemble methods with survival trees