# BOOSTING FOR COMPETING RISKS DATA WITH CURE FRACTION

Nilotpal Sanyal
Department of Mathematical Sciences, University of Texas at El Paso,
500 W. University Ave., El Paso, TX 79968, USA
Email: nsanyal@utep.edu

## Abstract

*Background.* Competing risk analysis is an important and active research area of survival analysis that deals with scenarios where individuals may experience multiple mutually exclusive events and thus, are subject to multiple 'competing' risks. In a population, there may also be individuals who can be considered, in some sense, 'cured' with respect to one or more of these competing events. Models that explicitly account for the proportion of cured individuals present in the population are called cure models. Competing risk analysis in cure models is a budding area with only a handful of works present in the literature. None of these works provide an efficient selection of predictive variables with high-dimensional sparse data.

*Proposed method.* A mixture cure model expresses the overall survival function as a mixture of survival functions for the cured and uncured individuals. We propose a novel method for selecting variables in a mixture cure model with competing risks based on boosting. Within a standard mixture cure model, we consider the standard mixture model representation of the competing risks. The resulting competing risk mixture cure model with selected variables is fitted using an expectation-maximization (EM)–type algorithm that alternates between performing an expectation step and a set of boosting steps. The inference is performed separately for the incidence part, which determines whether an event can occur or not, and the latency part, which determines the time of occurrence of an event given that an event can occur. We consider a component-wise boosting approach where in each boosting step, a single component of the parameter vector is updated using a Newton-Raphson step based on a penalized likelihood. The component that has the highest value of the score statistic is the one that is updated. Successive boosting steps are performed in this manner until a boosting convergence criterion is met. Finally, the algorithm is taken to converge when an expectation convergence criterion is met. Our method allows for the presence of mandatory variables that are always present in the model and are not subjected to penalization.

Our algorithm has four tuning parameters—the boosting convergence parameter, the expectation convergence parameter, and penalty parameters for the incidence and latency parts. Currently, we are working on developing suitable strategies to set their values based on objective or predictive considerations. We plan to apply our method to multiple high-dimensional datasets from epidemiology and public health.

**Keywords:** Competing risk analysis; cure model; boosting; expectation-maximization.