Dan Nordman Seminar

May 3, 2021

Title: Within-sample prediction of a number of future events

Abstract:

The talk overviews a prediction problem encountered in reliability engineering, where a need arises to predict the number of future events (e.g., failures) among a cohort of units associated with a time-to-event process. Examples include the prediction of warranty returns or the prediction of the number of product failures that could cause serious harm. The data consist of a collection of units observed over time where, at some freeze point in time, inspection then ends, leaving some units with realized event times while other units have not yet experienced events. In other words, data are right-censored, where (crudely expressed) some units have "died" by the freeze time while other units continue to "survive." The problem becomes predicting how many of the "surviving" units will have events (i.e., "death") in a next future interval of time, as quantified in terms of a prediction bound or interval. Because all units belong to the same data set, either by providing direct information (i.e., observed event times) or by becoming the subject of prediction (i.e., censored event times), such predictions are called within-sample predictions and differ from other prediction problems considered in most literature. A standard plug-in (also known as estimative) prediction approach turns out to be invalid for this problem (i.e., for even large amounts of data, the method fails to have correct coverage probability). However, several bootstrap-based methods can provide valid prediction intervals. To this end, a commonly used prediction calibration method is shown to be asymptotically correct for within-sample predictions, and two alternative predictive-distribution-based methods are presented that perform better than the calibration method.