Professor Tilmann Gneiting

Editor-in-Chief, Annals of Applied Statistics

Dear Professor Gneiting:

On behalf of myself and my co-authors, I would like to submit our paper “A Hierarchical Failure-Time Model for Observational Data Exhibiting Infant-Mortality and Wearout Failure Modes” for consideration for publication in the Annals of Applied Statistics.

This work began as a class project for Dr. Meeker’s course on Reliability. We found the Backblaze data set and sought to model the lifetime for each drive-model. Since the data were heavily censored and left-truncated, we knew we would use a likelihood-based approach with a parametric model. As others have noted elsewhere, exponential distributions did not fit the data. We chose to use Weibull distributions and model the parameters hierarchically since there were not enough data to fit many of the drive-models independently. At the end of the semester, Dr. Meeker encouraged us to develop the project into a journal article because hierarchical modeling is an approach not widely implemented in Reliability.

Careful evaluation of the results showed that the Weibull was inadequate for several drives, since there was evidence of infant mortality. Dr. Meeker suggested that we consider the Generalized Limited Failure Population (GLFP) model to handle it. While more somewhat more difficult to work with, we found that it not only provided a more realistic model for hard-drive lifetimes, but that our hierarchical approach enabled us to model lifetime using the GLFP even when the data for a particular drive-model were extremely limited.

The motivation for this work initially came from a publicly available data set with hard drive failures grouped by brand. There was little information in the data due to left truncation and heavy right censoring—hard disk drives rarely fail. Moreover, standard lifetime models were too simplistic to capture the bathtub hazard we observed in the data. As we developed our model we realized this problem of limited information, and multiple failure modes, is common in consumer product reliability data. Standard parametric models are too simple to capture multiple failure modes, and more complex models often require more information for estimation than is available. The hierarchical Generalized Limited Failure Population model we present in this paper addresses both of these problems in a unified and computationally straightforward framework, useful for practitioners.

I presented this work at JSM this past year and researchers from companies such as Gortex and Regeneron expressed strong interest in the model. Both companies have data that rarely fails making it difficult to fit realistic failure time models when to comparing lifetime distributions across products.

We hope that you and the reviewers like the paper. We will look forward to hearing from you at the completion of the review process.

Sincerely,