



Computational Graphics and Statistical Analysis: Mixed Type Random Variables, Confidence Regions, and Golden Quantile Rank Sets

Christopher Weld

College of William & Mary, Department of Applied Science, 2019

Field: Applied Mathematics Degree: Ph.D.

Advisor: Lawrence Leemis, Professor of Mathematics

Abstract

Mixed type random variables are a hybrid of continuous and discrete random variables, having components of both continuous probability density and discrete probability mass. This dissertation illustrates the challenges inherent in plotting mixed type distributions, and introduces an algorithm that addresses those issues. It considers sums and products of mixed type random variables, and supports its conclusions using Monte Carlo simulation experiments. Lastly, it introduces MixedAPPL, a computer algebra system software package designed for manipulating mixed type random variables.

Confidence regions are a multi-dimensional version of a confidence interval. They are helpful to visualize and quantify uncertainty surrounding a point estimate. We begin by developing efficient plot algorithms for two-dimensional confidence regions. This research focuses specifically on likelihood-ratio based confidence regions for two-parameter univariate probability models, although the plot techniques are transferable to any two-dimensional setting. The R package 'conf' is introduced, which automates these confidence region plot algorithms for complete and right-censored data sets. Among its benefits, 'conf' provides access to Monte Carlo simulation experiments for confidence region coverage to an extent not possible previously. The corresponding coverage analysis results include reference tables for the Weibull, normal, and log-logistic distributions. These reference tables yield confidence region plots with exact coverage.

The final topic is the introduction and analysis of a golden quantile rank set (GQRS). The term quantile rank set is used here to denote the population cumulative distribution function values corresponding to a sample. A GQRS can be thought of as "perfectly" representative of their population distribution because samples corresponding to a GQRS result in an estimator(s) matching the associated true population parameter(s). This unique characteristic is not applicable for all estimators and/or distributions, but when present, provides valuable insights and applications. Specifically, applications include an alternative (and at times computationally superior) method for parameter estimation and an exact actual coverage methodology for confidence regions (at times in which currently only estimates exist). Distributions with a GQRS associated with maximum likelihood estimation include the normal, exponential, Weibull, log logistic, and one-parameter exponential power distributions.