Handling the Elephant in the Room: Strategies to Address Model Uncertainty in Quantitative Analyses

S. Ferson

Centre de Recherches de Royallieu, Université de Technologie de Compiègne, F-60205 Compiègne Cedex, France, scott.ferson@utc.fr, sandp8@gmail.com

Keywords: *model uncertainty; structural uncertainty; polynomial regression; risk analysis.*

Abstract

How can we project uncertainty about X (which may be aleatory, epistemic or both) through a function f to characterize the uncertainty about $Y = f(X)$ when f itself has not been precisely characterized? Although the general problem of how to quantitatively express and project model uncertainty though mathematical calculations in a risk analysis can be addressed by only a few strategies, all of which seem either dubious or quite crude, there are a variety of special cases where methods to handle model uncertainty are rather well developed and available solutions are both comprehensive and subtle. For instance, uncertainty about the shapes of probability distributions can be captured as credal sets or p-boxes. Likewise, uncertainty about the stochastic dependencies between distributions can be projected using Kolmogorov–Fréchet bounding. Numerical experiments suggest that there is also another special case of model uncertainty that can be addressed fairly robustly: when evidence of the statistical relationship between variables has been condensed into regression analyses.

Suppose that the function f that has been characterized from sparse sample data (X_i,Y_i) via statistical regression. A straightforward convolution using regression statistics allows us to reconstruct the scatter of points processed in the original regression model, but it is well known that regression analysis does not necessarily select a model that actually reflects how data were generated. What if we do not know which order polynomial should have been used in the regression analysis? It turns out that the default reconstruction has conservative characteristics no matter what polynomial actually generated the data. We observe three facts: (1) models of all orders yield conservative characterizations of the variance of Y, (2) models of all orders yield reasonably conservative characterizations of the tail risks of Y, and (3) the envelope of resulting predicted Y distributions expressed as a p-box is conservative in all respects (i.e., it is effectively sure to enclose the real distribution of Y). This p-box seems to represent the model uncertainty induced in Y owing to the underlying uncertainty about X, *and* the model uncertainty about which degree polynomial is correct, contingent on the presumption that a polynomial model of some order is appropriate. These observations suggest a very simple and inexpensive strategy for computing conservative bounds on Y. Moreover, these expressions of uncertainty do not seem overly conservative. It does not matter which order a prior regression analysis may have used, so it is possible to obtain appropriately conservative estimates of tail risks for $Y = f(X)$ whatever model was used.