

Speaker:

Benjamin Kedem, Department of Mathematics & Institute for Systems Research, University of Maryland, College Park

Title:

Estimation of Small Tail Probabilities by Repeated out of Sample Fusion.

Abstract:

Often, it is required to estimate the small tail probability p that a quantity such as toxicity level, lead in dust, mercury in fish, chlorophenol in drinking water, methane exposure, plutonium contamination, etc., exceeds an unsafe high threshold. To estimate such a probability, information is needed about large values of the quantity of interest. However, in many cases, the data only contain values below or even far below the designated threshold, which makes the estimation of p challenging.

We shall present a novel statistical method whereby, under a mild assumption, it is possible to “peek” into the hidden domain above the threshold, and estimate p by repeated fusion of the given sample with externally generated random data. The consequential estimates based on moderately large samples are surprisingly precise. A comparison with peaks-over-threshold (POT) points to the merit of repeated out of sample fusion.

The gist of the method is an iterative procedure for capturing p along a certain curve referred to as a *B-Curve*. Thus, there is a curve. The curve contains a point “•”. We are after that point. The point gives p .

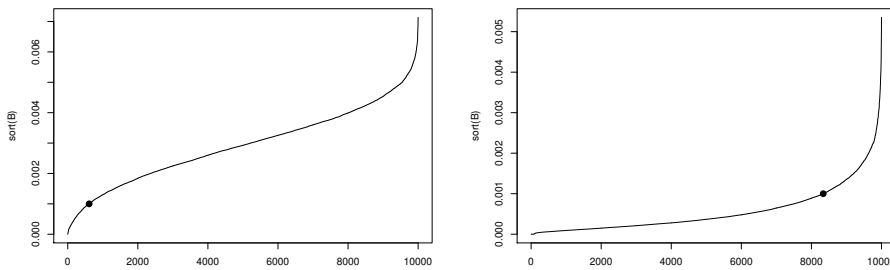


Figure 1: Typical B-Curves.