Simon Wilson

Using Storm for scaleable sequential statistical inference

In sequential statistical inference, data arrive as a stream and inference is an iterative process that updates as new data are available. Numerous examples and applications exist, starting with the Kalman filter and its generalisations such as the dynamic state space model. Approaches to implement inference in this setting are the subject of much current work e.g. sequential Monte Carlo. The challenge is not only to work with data sources that require sophisticated analyses, but also for scaleable inference algorithms that can cope with increased data dimension and arrival rates.

In this talk we describe a programming environment called Storm. This is one of several such environments for the processing of streaming data in a distributed manner. It is illustrated with 2 examples: computation of running summary statistics and a grid-based posterior approximation. The performance of these algorithms is evaluated and discussed with respect to these examples. Then some more recent work on the implementation of the ensemble Kalman filter is discussed. This is simultaneously a more useful method for complex analysis yet presents some difficulties for these streaming systems that seem to be common to many sequential statistical algorithms. We will report progress with our implementation.