Computation and Aggregation of Quantiles from Data Streams

John M. Chambers David A. James Diane Lambert
Scott Vander Wiel

We describe the Incremental Quantile (IQ) method and its R implementation. IQ is a quick-and-dirty quantile tracker that we developed for monitoring network applications, and in particular it addresses two problems: (1) how to estimate, under strict computational requirements, multiple quantiles on a very large number of data streams or "agents" (e.g., groups of IP/port/time combinations), and (2) how to combine these agent quantiles to estimate quantiles of arbitrary data aggregates without re-processing the raw data stream.

The IQ method is attractive for its simplicity: As new data become available, IQ updates the current quantile estimates from the empirical CDF (eCDF) obtained by (weighted) averaging the current quantile estimates' eCDF and the new data eCDF. On the other hand, careful selection of IQ's probabilities and buffer sizes are needed for adequate performance.

We close by showing that IQ outperforms the previously implemented algorithm, but we also show through simulations the high price we pay in terms of root mean squared error for the scant use of computing resources.