Abstract: This work deals with streaming algorithms for estimation of ranks and quantiles that perform a single pass through the input data stream using a small space. After reading a stream of N elements of a totally ordered universe, a streaming algorithm for rank (or quantile) estimation answers rank (or quantile) queries with *additive error* if the error is at most $\pm\varepsilon N$ and with *relative error* if for item y with rank $\mathrm{R}(y)$, the error is at most $\pm\varepsilon\,\mathrm{R}(y)$. The first problem is optimally solved by the KLL algorithm in space $\mathcal{O}(\varepsilon^{-1})$, and the best-known algorithm for the relative error is ReqSketch, which takes space $\mathcal{O}(\varepsilon^{-1}\log^{1.5} N)$.

Our algorithm called Jagged Sketch consists of two significant improvements to the ReqSketch algorithm. The first of the improvements reduces the error for high ranks by a factor of $\sqrt{\log(N)}$, the second one improves the error by a factor up to $\log(N)$ for important ranks chosen by the user and for ranks close to them, all while maintaining the same space complexity. We support our theoretical analysis by experiments that demonstrate that Jagged Sketch can indeed reduce the error for selected ranks while maintaining the same space and similar error for other ranks compared to ReqSketch.

For $\varepsilon \in \mathcal{O}(\log^{-1.5} N)$ Jagged Sketch achieves additive error in the same space as KLL while simultaneously retaining near-relative error guarantee. In practice, the error for large ranks is about four times larger, while for the lowest ranks, it is up to about a hundred times smaller.