Quantile Sketch Algorithms

Levente Birszki

Supervisors:

Dr. Gábor Rétvári
Dr. Balázs Vass

Outline

- Basic concepts
- Approximating quantiles
- Results
- Future plans

Basic concepts

Definition (sketch)

A *sketch* $S(X)$ of some data set X with respect to some function f is a compression of X that allows us to compute, or approximately compute $f(X)$ given access only to $S(X)$.

Definition (rank)

Given an x element from the input stream. $r(x)$, the *rank* of x is the number of elements smaller or equal than x in the sorted input.

Definition (quantile)

The q-quantile for $q \in [0, 1]$ is the element x_q, whose rank is $\lceil qn \rceil$.

Error guaranties

Definition (rank error)

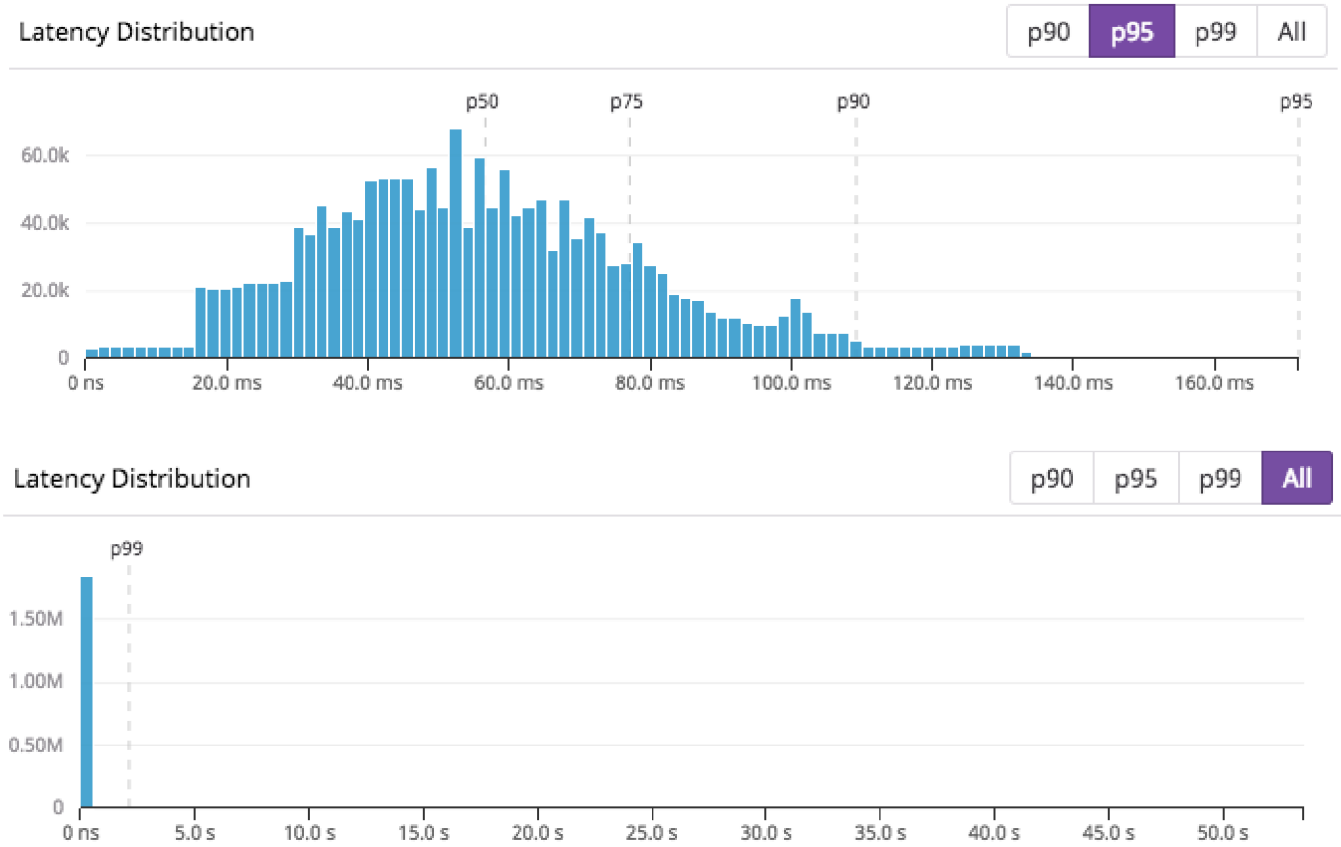
An element $\tilde{x}_q$ is an ε-approximate q-quantile if $|r(x_q) - r(\tilde{x}_q)| \leq \varepsilon n$. This also known as rank error.

Definition (relative error)

$\tilde{x}_q$ is an α-accurate q-quantile if $|x_q - \tilde{x}_q| \leq \alpha x_q$, for a given x_q. This is called relative error.

Rank error vs relative error

Figure: Histograms for $p0$-$p95$ and $p0$-$p100$ of 2 million web request response times.



Results

Definition (single quantile approximation problem)

In the *single quantile approximation problem*, given an $x_1, \ldots, x_n$ input stream, q, ε and δ. Construct a streaming algorithm, which computes an ε-approximate q-quantile with probability at least $1 - \delta$.

Publication	Algorithm	Space Complexity	Mergeability	quantile type
2001	GK-sketch	$O\left(\frac{1}{\varepsilon}\log(\varepsilon n)\right)$	no	all
2004	q-digest	$O(\frac{1}{\varepsilon}\log u)$	yes	all
2016	KLL	$O(\frac{1}{\varepsilon}\log^2\log\frac{1}{\delta})$	yes	singe
2016	KLL	$O(\frac{1}{\varepsilon}\log^2\log\frac{1}{\delta\varepsilon}))$	yes	all
2017	FO	$O(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon})$	no	all
2019	SweepKLL	$O(\frac{1}{\varepsilon}\log\log\frac{1}{\delta}))$	no	single
2019	SweepKLL	$O(\frac{1}{\varepsilon}\log\log\frac{1}{\delta\varepsilon}))$	no	all

Future plans

- Examine relative error sketching algorithms such as DDSketch, and ReqSketch.
- With a quantile sketch, we can approximate the CDF of the input stream. We could use this information to improve the algorithm.
- Creating a sketching algorithm for finding a few, predefined quantiles, using the previous idea.

Thank you for your attention!