

Abstrakt: Tato práce se zabývá proudovými (streaming) algoritmy pro odhad distribucí a kvantilů, které provedou jeden průchod přes vstupní proud dat za použití malé paměti. Po načtení proudu N prvků z lineárně uspořádaného univerza proudový algoritmus pro odhad kvantilů odpovídá s *aditivní chybou*, když je velikost chyby nejvýše $\pm \varepsilon N$ a s *relativní chybou*, když je pro item y s rankem $R(y)$ chyba nejvýše $\pm \varepsilon R(y)$. První z těchto problémů řeší optimálně algoritmus KLL v prostoru ε^{-1} , nejlepší známý algoritmus pro relativní chybu je ReqSketch, který potřebuje prostor $\varepsilon^{-1} \log^{1.5}(N)$.

Náš algoritmus Jagged Sketch spočívá ve dvou vylepšeních algoritmu ReqSketch. První z vylepšení zmenšuje chybu $\sqrt{\log(N)}$ -krát pro vysoké ranky, druhé zmenšuje chybu až $\log(N)$ -krát pro důležitý rank dle volby uživatele a ranky jemu blízké, to vše při zachování stejného prostoru. Teoretickou analýzu jsme podpořili experimenty, které prokazují, že Jagged Sketch dokáže oproti ReqSketch skutečně snížit chybu pro vybrané ranky při zachování stejného prostoru a podobné chyby pro ostatní ranky.

Pro $\varepsilon \in \mathcal{O}(\log^{-1.5} N)$ odpovídá Jagged Sketch s aditivní chybou ve stejném prostoru jako KLL, přičemž si zároveň zachovává garanci téměř relativní chyby. V praxi je chyba Jagged Sketche pro velké ranky přibližně čtyřikrát větší, zatímco pro malé ranky je až stokrát menší.