

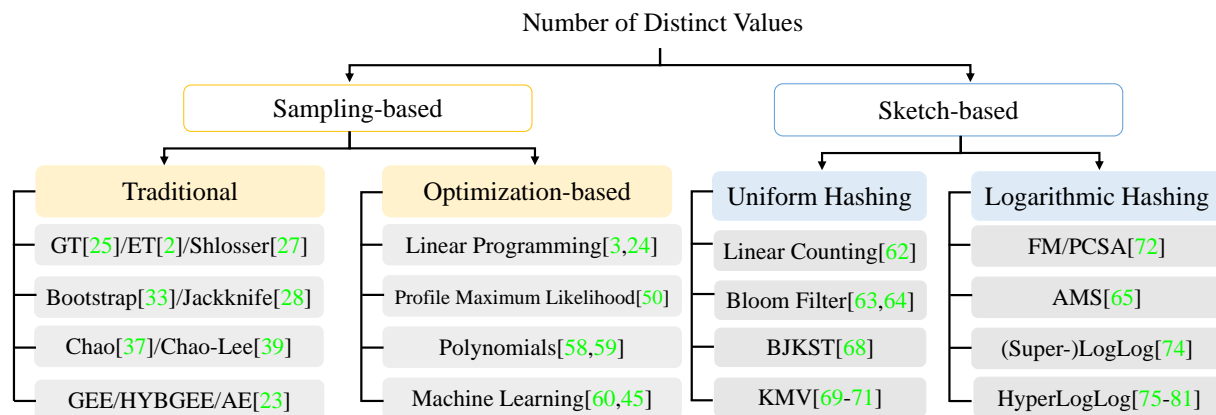
A survey of estimating number of distinct values

Jiajun Li, Runlin Lei, Zhewei WEI

Frontiers of Computer Science, DOI: [10.1007/s11704-024-40952-3](https://doi.org/10.1007/s11704-024-40952-3)

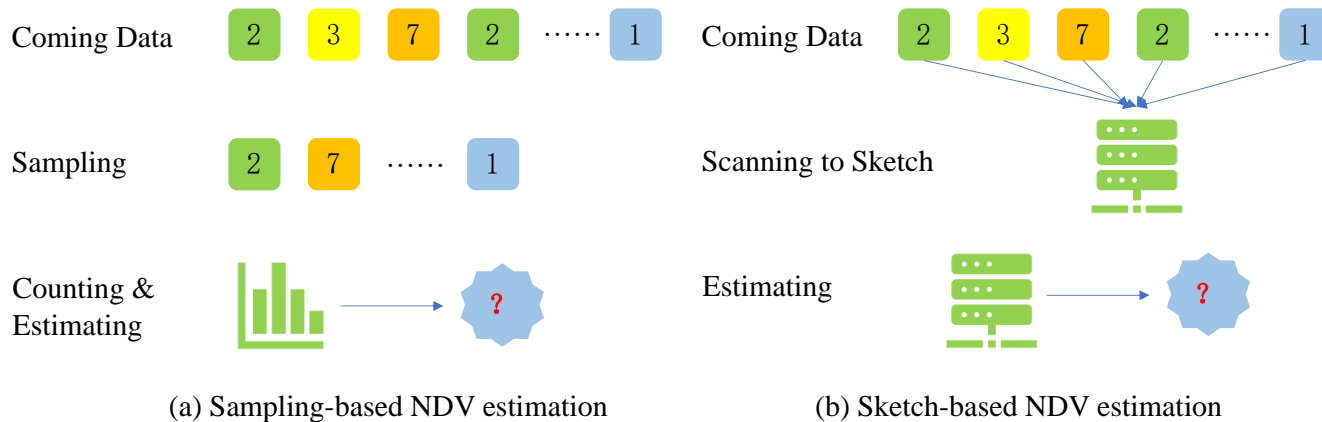
Background & Ideas

- Background of **estimating number of distinct values**:
 - Estimating the Number of Distinct Values (NDV) is a critical task in the fields of databases and data streams and various algorithms for estimating NDV has been developed.
 - The main NDV estimation methods have been categorized into two groups, but as complex environments change, NDV estimation algorithms face new opportunities and challenges.
- Ideas: This survey examines **the latest algorithms for NDV estimation**, delves into the strength and weakness of different algorithms, alongside an exploration of current research directions and future prospects.



Main Contributions

- Contributions:
 - A comprehensive review of **sampling- and sketch-based NDV estimation algorithm** and we analyze in depth their advantages and disadvantages;
 - We present in detail the technical improvements made to existing NDV estimation algorithms to adapt them to the current more complex environments.
 - We discuss the **challenges** this field faces and point out possible future research directions and technology trends.



Two types of algorithms for NDV estimation.