

**Chambers, John M.; James, David A.; Lambert, Diane; Vander Wiel, Scott**  
**Monitoring networked applications with incremental quantile estimation.** (English)

Zbl 1426.62379

Stat. Sci. 21, No. 4, 463-475 (2006).

Summary: Networked applications have software components that reside on different computers. Email, for example, has database, processing, and user interface components that can be distributed across a network and shared by users in different locations or work groups. End-to-end performance and reliability metrics describe the software quality experienced by these groups of users, taking into account all the software components in the pipeline. Each user produces only some of the data needed to understand the quality of the application for the group, so group performance metrics are obtained by combining summary statistics that each end computer periodically (and automatically) sends to a central server. The group quality metrics usually focus on medians and tail quantiles rather than on averages. Distributed quantile estimation is challenging, though, especially when passing large amounts of data around the network solely to compute quality metrics is undesirable. This paper describes an Incremental Quantile (IQ) estimation method that is designed for performance monitoring at arbitrary levels of network aggregation and time resolution when only a limited amount of data can be transferred. Applications to both real and simulated data are provided.

**MSC:**

62P30 Applications of statistics in engineering and industry; control charts

Cited in **1** Review  
Cited in **2** Documents

**Keywords:**

aggregated data; data stream; performance monitoring; reliability

**Full Text:** [DOI](#) [Euclid](#)

**References:**

- [1] Chen, F., Lambert, D. and Pinheiro, J. C. (2000). Incremental quantile estimation for massive tracking. In Proc. Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 516–522. ACM Press, New York.
- [2] Greenwald, M. B. and Khanna, S. (2001). Space-efficient online computation of quantile summaries. In Proc. 2001 ACM SIGMOD International Conference on Management of Data 58–66. ACM Press, New York.
- [3] Greenwald, M. B. and Khanna, S. (2004). Power-conserving computation of order-statistics over sensor networks. In Proc. 23rd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS 2004) 275–285. ACM Press, New York.
- [4] Liechty, J. C., Lin, D. K. J. and McDermott, J. P. (2003). Single-pass low-storage arbitrary quantile estimation for massive datasets. Stat. Comput. 13 91–100. · doi:10.1023/A:1023296123228
- [5] Manku, G. S., Rajagopalan, S. and Lindsay, B. G. (1998). Approximate medians and other quantiles in one pass and with limited memory. In Proc. 1998 ACM SIGMOD International Conference on Management of Data 426–435. ACM Press, New York.
- [6] McDermott, J. P., Babu, G., Liechty, J. C. and Lin, D. K. J. (2003). Data skeletons: Simultaneous estimation of multiple quantiles for massive streaming datasets with applications to density estimation. Technical Report #03-02, Dept. Statistics, Pennsylvania State Univ.
- [7] Munro, J. and Paterson, M. (1980). Selection and sorting with limited storage. Theoret. Comput. Sci. 12 315–323. · Zbl 0441.68067 · doi:10.1016/0304-3975(80)90061-4
- [8] Robbins, H. and Monro, S. (1951). A stochastic approximation method. Ann. Math. Statist. 22 400–407. · Zbl 0054.05901 · doi:10.1214/aoms/1177729586 ·
- [9] Tierney, L. (1983). A space-efficient recursive procedure for estimating a quantile of an unknown distribution. SIAM J. Sci. Statist. Comput. 4 706–711. · Zbl 0524.65099 · doi:10.1137/0904048

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.