Bowen, Claire McKay; Liu, Fang; Su, Bingyue
Differentially private data release via statistical election to partition sequentially. (English)

Summary: Differential Privacy (DP) formalizes privacy in mathematical terms and provides a robust concept for privacy protection. Differentially Private Data Synthesis (DIPS) techniques produce and release synthetic individual-level data in the DP framework. One key challenge to develop DIPS methods is the preservation of the statistical utility of synthetic data, especially in high-dimensional settings. We propose a new DIPS approach, STatistical Election to Partition Sequentially (STEPS) that partitions data by attributes according to their importance ranks according to either a practical or statistical importance measure. STEPS aims to achieve better original information preservation for the attributes with higher importance ranks and produce thus more useful synthetic data overall. We present an algorithm to implement the STEPS procedure and employ the privacy budget composability to ensure the overall privacy cost is controlled at the pre-specified value. We apply the STEPS procedure to both simulated data and the 2000–2012 Current Population Survey youth voter data. The results suggest STEPS can better preserve the population-level information and the original information for some analyses compared to PrivBayes, a modified Uniform histogram approach, and the flat Laplace sanitizer.

MSC:
62P30 Applications of statistics in engineering and industry; control charts
68P27 Privacy of data

Keywords: privacy budget; differentially private data synthesis (DIPS); general utility; propensity score; universal histogram

Full Text: DOI arXiv

References:


Li, KH; Meng, XL; Raghunathan, TE; Rubin, DB, Significance levels from repeated p-values with multiply-imputed data, Stat. Sin., 1, 65-92 (1991) · Zbl 0823.62009


McDonald, MP; Thornburg, M., Registering the youth through voter preregistration, NYUJ Legis Pub Pol’y, 13, 551 (2010)


Qardaji, W., Yang, W., Li, N.: Differentially private grids for geospatial data. In: Proceedings of 2013 IEEE 29th International Conference on Data Engineering (ICDE), pp. 757-768 (2013)


doi:10.1111/rssa.12358
doi:10.1198/jasa.2009.tm08651
doi:10.1007/978-3-642-15546-8\textbackslash _1

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.