

Privacy protection and quantile estimation under noise multiplication

Bimal Sinha¹, Tapan Nayak², Laura Zayatz³

¹ *University of Maryland Baltimore County and US Census Bureau*
email: sinha@umbc.edu

² *George Washington University and US Census Bureau*, email: tapan@gwu.edu

³ *US Census Bureau*, email: laura.zayatz@census.gov

Keywords: confidentiality, disclosure risk, Metropolis algorithm, noise multiplication, prior distribution.

We will address two inferential aspects of noise multiplied magnitude microdata. First, in the context of disclosure risk assessment of tabular magnitude data, we study the consequences of noise multiplication of original microdata when an intruder tries to speculate a target unit's value in a cell based on knowledge of the noise perturbed cell total and values of some original units within the cell. Second, we discuss statistical methods to infer about a *quantile* of a microdata set based on their noise perturbed values. An application with income data will be presented.

References

- [1] Blumenthal, S., Cohen, A. (1968). Estimation of the larger translation parameter. *Ann. Math. Statist.* **39**, 502–516.
- [2] Blumenthal, S., Cohen, A. (1968). Estimation of the larger of two normal means. *Ann. Math. Statist.* **39**, 861–876.
- [3] Brand, R. (2002). Microdata protection through noise addition. In: *Inference Control in Statistical Databases*, Ed. J. Domingo-Ferrer. Berlin: Springer, 97–116.
- [4] Doyle, P., Lane, J., Theeuwes, J., Zayatz, L. (Ed.) (2001). *Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies*. Amsterdam: Elsevier.
- [5] Dudewicz, E.J. (1971). Maximum likelihood estimators for non-1-1 functions. *Trabajos de Estadística y de Investigación Operativa* **22**, 65–70.
- [6] Dudewicz, E.J. (1971). Maximum likelihood estimators for ranked means. *Z. Wahrscheinlichkeitstheorie Verw. Grb.* **19**, 29-42.
- [7] Duncan, G.T., Fienberg, S.E. (1999). Obtaining information while preserving privacy: A Markov perturbation method for tabular data. In: *Eurostat Statistical Data Protection '98 Lisbon*, Luxembourg: Eurostat, 351–362.
- [8] Duncan, G.T., Stokes, S.L. (2004). Disclosure risk vs. data utility: The R-U confidentiality map as applied to topcoding. *Chance* **17**, 16–20.
- [9] Duncan, G. T., Mukherjee, S. (2000). Optimal disclosure limitation strategy in statistical databases: Detering tracker attacks through additive noise. *J. Amer. Statist. Assoc.* **95**, 720–729.
- [10] Elfessi, A., Pal, N. (1992). Estimation of the smaller and larger of two uniform scale parameters. *Commun. Statist. - Theory Meth.* **21**, 2997–3015.

- [11] Evans, T., Zayatz, L., Slanta, J. (1998). Using noise for disclosure limitation of establishment tabular data. *J. Official Statist.* **4**, 537-551.
- [12] Fong, D.K.H. (1987). *Ranking and Estimation of Exchangeable Means in Balanced and Unbalanced Models: A Bayesian Approach*. Ph.D. thesis, Purdue University.
- [13] Fong, D.K.H. (1992). A Bayesian approach to the estimation of the largest normal mean. *J. Statist. Comput. Simul.* **40**, 119-133.
- [14] Fuller, W.A. (1993). Masking procedures for microdata disclosure limitation. *J. Official Statist.*, 383-406.
- [15] Givens, G.H., Hoeting, J.A. (2005). *Computational Statistics*. New York: John Wiley.
- [16] Karr, A.F., Kohnen, C.N., Oganian, A., Reiter, J.P., Sanil, A.P. (2006). A framework for evaluating the utility of data altered to protect confidentiality. *Amer. Statist.* **60**, 224-232.
- [17] Kim, J. (1986). A method for limiting disclosure in microdata based on random noise and transformation. In: *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 303-308.
- [18] Kim, J.J., Winkler, W.E. (2003). Multiplicative noise for masking continuous data. Technical Report Statistics #2003-01, Statistical Research Division, U.S. Bureau of the Census, Washington D.C., April 2003.
- [19] Kumar, S., Sharma, D. (1993). Unbiased inestimability of ordered parameters. *Statistics* **24**, 137-142.
- [20] Little, R.J.A. (1993). Statistical analysis of masked data. *J. Official Statist.* **9**, 407-426.
- [21] Misra, N., Anand, R., Singh, H. (1997). Estimation of the smaller and larger scale parameters of two exponential distributions. *Statistics and Decisions* **15**, 75-98.
- [22] Nayak, T. Sinha, B., Zayatz, L. (2010). Statistical properties of multiplicative noise masking for confidentiality protection. To appear in *J. Official Statist.*
- [23] Shao, J. (1999). *Mathematical Statistics*. New York: Springer.
- [24] Tendick, P. (1991). Optimal noise addition for preserving confidentiality in multivariate data. *J. Statist. Plann. Inference* **27**, 341-353.
- [25] van Eeden, C. (2006). *Restricted Parameter Space Estimation Problems*. New York: Springer.
- [26] Willenborg, L.C.R.J., De Waal, T. (2001). *Elements of Statistical Disclosure Control*. New York: Springer.
- [27] Yancey, W.E., Winkler, W.E., Creecy, R.H. (2002). Disclosure risk assessment in perturbative microdata protection. In: *Inference Control in Statistical Databases*, Ed. J. Domingo-Ferrer, Springer, 135-152.