

ABSTRACT OF THE DISSERTATION

Masking Microdata Files with Mixtures of Multivariate Normal Distributions

by

Gina Marie Roque

Doctor of Philosophy, Graduate Program in Applied Statistics
University of California, Riverside, June, 2000
Professor S. James Press, Chairperson

Many methods for masking microdata have been introduced in the past 25 years. Kim's (1986) introduction of the additive noise method for limiting disclosure appeared most promising because it preserves many statistical properties. It fails, however, to preserve confidentiality for a sizable percentage of respondents (Winkler 1998). Other schemes sacrifice the preservation of statistical properties (or the scope of potential analyses) for a high level of confidentiality.

We consider a generalization of Kim's method and generate noise from a finite mixture of multivariate normal distributions. Sequential Quadratic Programming is used to obtain values for the parameters and to control the amount of noise added to observations from each attribute. We present several examples to demonstrate the proposed masking scheme preserves statistical properties as well as confidentiality. We measure the risk of disclosure with the application of Record Linkage Software.