

Working Paper No. 5 (Summary)

ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)**

Joint ECE/Eurostat work session on statistical data confidentiality
(Luxembourg, 7-9 April 2003)

Topic (i): New theories and emerging methods

**BAYESIAN NETWORKS REPRESENTATIONS OF CONTINGENCY TABLES
FOR REDUCING DISCLOSURE AND
PRESERVING THE ACCURACY OF SUFFICIENT STATISTICS**

Invited Paper

Submitted by the U.S. Bureau of the Census, United States¹

¹ Prepared by Bill Winkler (william.e.winkler@census.gov) and Yves Thibaudeau (yves.thibaudeau@census.gov).

Bayesian Networks Representations of Contingency Tables for Reducing Disclosure and Preserving the Accuracy of Sufficient Statistics

Yves Thibaudeau and William E. Winkler

The paper explores the use of Bayesian networks to simulate discrete data reported in a contingency table in a manner that reconstructs the contingency table with reduced risk of disclosure, relative to the original table. The object of the investigation is the degree of integrity of the network in maintaining the accuracy of certain sufficient statistics through the simulation process. While with general log-linear models it is often possible to guarantee a realistic simulation of the interactions between specific variables, the parsimony of Bayesian networks implies a reduced level of control in reproducing interactions. We experiment with different techniques to construct Bayesian networks and compute the associated conditional probabilities in order to simulate entire contingency tables, or selected cells in the tables. At the same time we measure the level of accuracy of certain sufficient statistics computed from the modified table, relative to the original table. We attempt to delineate a strategy for identifying Bayesian networks for simulating contingency tables that maintain a relatively good accuracy for certain sufficient statistics, while also reducing the risk of disclosure inherent to the original table.