On the Security of Noise Addition for Privacy in Statistical Databases

Josep Domingo-Ferrer, Francesc Sebé, and Jordi Castellà-Roca

Universitat Rovira i Virgili, Dept. of Computer Engineering and Mathematics
Av. Països Catalans 26, E-43007 Tarragona, Catalonia
{jdomingo,fsebe,jcaste}@etse.urv.es

Abstract. Noise addition is a family of methods used in the protection of the privacy of individual data (microdata) in statistical databases. This paper is a critical analysis of the security of the methods in that family.

Keywords: Noise addition, Statistical database protection, Statistical disclosure control, Statistical disclosure limitation, Data security, Respondents' privacy.

1 Introduction

Privacy in statistical databases is about finding tradeoffs to the tension between the increasing societal and economical demand for accurate information and the legal and ethical obligation to protect the privacy of individuals and enterprises which are the source of the statistical data. To put it bluntly, statistical agencies cannot expect to collect accurate information from individual or corporate respondents unless these feel the privacy of their responses is guaranteed; also, surveys of web users [6, 7, 11] show that a majority of these are unwilling to provide data to a web site unless they know that privacy protection measures are in place.

To achieve privacy without losing accuracy, statistical disclosure control (SDC) methods must be applied to data before they are released [16]. Data in statistical databases are of two kinds: tables (aggregate data) data and microdata (individual respondent data). SDC methods for microdata are also known as masking methods. Example masking methods include noise addition, resampling, blanking, imputation, microaggregation, etc. (see [5] for a survey).

1.1 Contribution and Plan of This Paper

This paper concentrates on masking methods based on noise addition and analyzes their security. It will be shown that the distribution of the original data can be reconstructed from the masked data, which in some cases may lead to disclosure. In fact, another critique to the security of noise addition has recently been published for the univariate case ([8]); using a different approach, we show that multivariate noise addition is not free from problems either.

J. Domingo-Ferrer and V. Torra (Eds.): PSD 2004, LNCS 3050, pp. 149–161, 2004.

Section 2 reviews noise addition methods for privacy protection. Section 3 presents a procedure for recontructing the original distribution of the original multivariate dataset given the masked data set and the noise distribution. Section 4 shows how the previous procedure can be applied to attack most practical noise addition methods. Empirical results of attacks are presented in Section 5. Section 6 is a conclusion. The Appendix contains the mathematical derivation of the reconstruction procedure given in Section 3.

2 Noise Addition Methods for Privacy Protection

We will sketch in this section the operation of the main additive noise algorithms in the literature. For a more comprehensive description and a list of references, see the excellent survey [3].

2.1 Masking by Uncorrelated Noise Addition

Masking by additive noise assumes that the vector of observations x_j for the j-th variable of the original data set X_j is replaced by a vector

$$z_j = x_j + \epsilon_j \tag{1}$$

where ϵ_j is a vector of normally distributed errors drawn from a random variable $\varepsilon_j \sim N(0, \sigma^2_{\varepsilon_j})$, such that $Cov(\varepsilon_t, \varepsilon_l) = 0$ for all $t \neq l$ (white noise).

The general assumption in the literature is that the variances of the ε_j are proportional to those of the original variables. Thus, if σ^2_j is the variance of X_j, then $\sigma^2_{\varepsilon_j} := \alpha\sigma^2_j$.

In the case of a p-dimensional data set, simple additive noise masking can be written in matrix notation as

$$Z = X + \varepsilon$$

where $X \sim (\mu, \Sigma), \varepsilon \sim N(0, \Sigma_\varepsilon)$ and

$$\Sigma_\varepsilon = \alpha \cdot diag(\sigma^2_1, \sigma^2_2, \cdots, \sigma^2_p), \text{ for } \alpha > 0$$

This method preserves means and covariances, *i.e.*

$$E(Z) = E(X) + E(\varepsilon) = E(X) = \mu$$

$$Cov(Z_j, Z_l) = Cov(X_j, X_l) \ \ \forall j \neq l$$

Unfortunately, neither variances nor correlation coefficients are preserved:

$$V(Z_j) = V(X_j) + \alpha V(X_j) = (1 + \alpha)V(X_j)$$

$$\rho_{Z_j, Z_l} = \frac{Cov(Z_j, Z_l)}{\sqrt{V(X_j)V(X_l)}} = \frac{1}{1+\alpha}\rho_{X_j, X_l} \ \ \forall j \neq l$$

2.2 Masking by Correlated Noise Addition

Correlated noise addition also preserves means and additionally allows preservation of correlation coefficients. The difference with the previous method is that the covariance matrix of the errors is now proportional to the covariance matrix of the original data, *i.e.* $\varepsilon \sim N(0, \Sigma_\varepsilon)$, where $\Sigma_\varepsilon = \alpha\Sigma$.

With this method, we have that the covariance matrix of the masked data is

$$\Sigma_Z = \Sigma + \alpha\Sigma = (1+\alpha)\Sigma \tag{2}$$

Preservation of correlation coefficients follows, since

$$\rho_{Z_j,Z_l} = \frac{1+\alpha}{1+\alpha}\frac{Cov(X_j,X_l)}{\sqrt{V(X_j)V(X_l)}} = \rho_{X_j,X_l}$$

Regarding variances and covariances, we can see from Equation (2) that masked data only provide biased estimates for them. However, it is shown in [10] that the covariance matrix of the original data can be consistently estimated from the masked data as long as α is known.

As a summary, masking by correlated noise addition outputs masked data with higher analytical validity than masking by uncorrelated noise addition. Consistent estimators for several important statistics can be obtained as long as α is revealed to the data user. However, simple noise addition as discussed in this section and in the previous one is seldom used because of the very low level of protection it provides [14, 15].

2.3 Masking by Noise Addition and Linear Transformations

In [9], a method is proposed that ensures by additional transformations that the sample covariance matrix of the masked variables is an unbiased estimator for the covariance matrix of the original variables. The idea is to use simple additive noise on the p original variables to obtain overlayed variables

$$Z_j = X_j + \varepsilon_j, \text{ for} j = 1, \cdots, p$$

As in the previous section on correlated masking, the covariances of the errors ε_j are taken proportional to those of the original variables. Usually, the distribution of errors is chosen to be normal or the distribution of the original variables, although in [12] mixtures of multivariate normal noise are proposed.

In a second step, every overlayed variable Z_j is transformed into a masked variable G_j as

$$G_j = cZ_j + d_j$$

In matrix notation, this yields

$$Z = X + \varepsilon$$

$$G = cZ + D = c(X + \varepsilon) + D$$

where $X \sim (\mu, \Sigma)$, $\varepsilon \sim (0, \alpha\Sigma)$, $G \sim (\mu, \Sigma)$ and D is a matrix whose j-th column contains the scalar d_j in all rows. Parameters c and d_j are determined under the restrictions that $E(G_j) = E(X_j)$ and $V(G_j) = V(X_j)$ for $j = 1, \cdots, p$. In fact, the first restriction implies that $d_j = (1 - c)E(X_j)$, so that the linear transformations depend on a single parameter c.

Due to the restrictions used to determine c, this methods preserves expected values and covariances of the original variables and is quite good in terms of analytical validity. Regarding analysis of regression estimates in subpopulations, it is shown in [10] that (masked) sample means and covariances are asymptotically biased estimates of the corresponding statistics on the original subpopulations. The magnitude of the bias depends on the parameter c, so that estimates can be adjusted by the data user as long as c is revealed to her.

The most prominent shortcomings of this method are that it does not preserve the univariate distributions of the original data and that it cannot be applied to discrete variables due to the structure of the transformations.

2.4 Masking by Noise Addition and Nonlinear Transformations

An algorithm combining simple additive noise and nonlinear transformation is proposed in [13]. The advantages of this proposal are that it can be applied to discrete variables and that univariate distributions are preserved.

The method consists of several steps:

1. Calculating the empirical distribution function for every original variable;
2. Smoothing the empirical distribution function;
3. Converting the smoothed empirical distribution function into a uniform random variable and this into a standard normal random variable;
4. Adding noise to the standard normal variable;
5. Back-transforming to values of the distribution function;
6. Back-transforming to the original scale.

In the European project CASC (IST-2000-25069), the practicality and usability of this algorithm was assessed. Unfortunately, the internal CASC report [4] concluded that

> All in all, the results indicate that an algorithm as complex as the one proposed by Sullivan can only be applied by experts. Every application is very time-consuming and requires expert knowledge on the data and the algorithm.

2.5 Section Summary

Thus, in practice, only simple noise addition or noise addition with linear transformation are used. When using linear transformations, the parameter c determining the transformations is revealed to the data user to allow for bias adjustment in the case of subpopulations.

3 Reconstructing the Original Distribution from Noise-Added Multivariate Data

In [1], an algorithm for reconstructing the original distribution of univariate data masked by noise addition is presented. In [2], a reconstruction algorithm with the same purpose is proposed which not only converges but does so to the maximum likelihood estimate of the original distribution. When comparing both proposals, it turns out that, even though the algorithm in [2] has nicer theoretical properties, the algorithm in [1] lends itself better to a multivariate generalization. Such a generalization is presented in this section.

We consider an original multivariate data set consisting of n records $x_1, x_2, \cdots, x_n$ with p dimensions each ($x_i = (x_i^1, \cdots, x_i^p)$). The n original records are modeled as realizations of n independent identically distributed random variables $X_1, X_2, \cdots, X_n$, each with the same distribution as a random variable X. To hide these records, n independent p-variate variables $Y_1, Y_2, \cdots, Y_n$ are used, each with the same distribution as a random variable Y. The published data Z set will be $z_1 = x_1 + y_1, z_2 = x_2 + y_2, \cdots, z_n = x_n + y_n$.

The purpose of our algorithm is to estimate the density function of the data set X (namely $f'_X(a^1, \cdots, a^p)$) from our knowledge of the published data Z and the density function f_Y.

Rather than estimating f'_X itself, the algorithm estimates the probability that X takes values in a certain interval. The technique is a multivariate generalization of the approach described in [1] for the univariate case ($p = 1$).

We first partition $\mathbb{R}^p$ into p-dimensional intervals (the i-th dimension is divided into k^i subintervals) and consider the grid formed by the midpoints of such intervals. Let $I_{a^1,\cdots,a^p}$ be the interval in which $(a^1, \cdots, a^p)$ lies, and let $m(I_{a^1,\cdots,a^p})$ be the midpoint of $I_{a^1,\cdots,a^p}$. The following approximations will be used:

- $f_Y(a^1, \cdots, a^p)$ will be approximated by $f_Y(m(I_{a^1,\cdots,a^p}))$
- $f_X(a^1, \cdots, a^p)$ will be approximated by the average of f_X over the interval in which $(a^1, \cdots, a^p)$ lies.

Let $N(I_{q^1,\cdots,q^p})$ be the number of points in Z that lie in interval $I_{q^1,\cdots,q^p}$, *i.e.* the number of elements in the set $\{(z_i^1, \cdots, z_i^p) | (z_i^1, \cdots, z_i^p) \in I_{q^1,\cdots,q^p}\}$

We can now state the reconstruction algorithm as follows (see Appendix for details on the mathematical derivation of the algorithm):

Algorithm 1 (p-dimensional reconstruction algorithm)

1. *Let $Pr^0(X \in I_{q^1,\cdots,q^p})$ be the prior probability that X takes a value in $I_{q^1,\cdots,q^p}$. For convenience, compute Pr^0 taking as prior distribution for X the p-dimensional uniform distribution over a p-dimensional interval in $\mathbb{R}$.*
2. *Let $j = 0$ (j is the iteration counter).*

3. Repeat
(a) $Pr^{j+1}(X \in I_{q^1,\cdots,q^p})$

$$= \frac{1}{n} \sum_{s^1=1}^{k^1} \cdots \sum_{s^p=1}^{k^p} N(I_{s^1,\cdots,s^p}) \times$$

$$\times \frac{f_Y(m(I_{s^1,\cdots,s^p}) - m(I_{q^1,\cdots,q^p}))Pr^j(X \in I_{q^1,\cdots,q^p})}{\sum_{t^1=1}^{k^1} \cdots \sum_{t^p=1}^{k^p} f_Y(m(I_{s^1,\cdots,s^p}) - m(I_{t^1,\cdots,t^p}))Pr^j(X \in I_{t^1,\cdots,t^p})}$$

(b) j:=j+1
until stopping criterion met.

4 Attacking Noise Addition Methods

As pointed out in Section 2.5 above, the noise addition approaches used in practice are limited to correlated noise addition and noise addition combined with linear transformations. Our attacks will be targeted at those two approaches.

4.1 Attacking Correlated Noise Addition

As noted in Section 2.2, for consistent estimators to be obtained from data masked with correlated noise addition, the parameter α must be revealed to the data user. If a user knows α, we have from Equation (2) that she can estimate the covariance matrix $\alpha\Sigma$ of the added noise as $\alpha\hat{\Sigma}$, where

$$\hat{\Sigma} = (1+\alpha)^{-1}\hat{\Sigma}_Z$$

and $\hat{\Sigma}_Z$ is the sample covariance matrix estimated from the masked data.

In this way, a user is able to estimate the distribution of the noise Y added when masking the data as a $N(0, \alpha\hat{\Sigma})$ distribution. Using the masked data set and the estimated noise distribution, the user is in a position to run Algorithm 1 to estimate the distribution of the original data X.

Now, if the user is malicious, she is interested in determining intervals $I_{q_1,\cdots,q_p}$ which are narrow in one or several dimensions and for which the number of records in the original data set that fall in the interval is one or close to one. Note that the number of original records $(x_1^i, \cdots, x_p^i)$ that lie in interval $I_{q_1,\cdots,q_p}$ can be estimated using the distribution of the original data as $NPr(X \in I_{q_1,\cdots,q_p})$, where N is the total number of individuals in the original data set. Such intervals disclose the value of one or more attributes for the individual(s) behind the record(s) lying in the interval.

4.2 Attacking Masking by Noise Addition and Linear Transformation

It was pointed out in Section 2.3 that the linear transformations used to complement correlated noise addition actually depend on a single parameter c, which must be revealed to the user if the latter is to adjust for biases in subpopulation estimates.

Now, a user can use her knowledge of c to estimate the independent term d_j of each linear transformation as

$$\hat{d}_j = (1 - c)\hat{E}(G_j)$$

where $\hat{E}(G_j)$ is the sample average of the j-th masked variable. Next, the user can undo the linear transformation to recover an estimate of the j-th overlayed variable Z_j:

$$\hat{Z}_j = c^{-1}(G_j - \hat{d}_j)$$

In this way, the values taken by all overlayed variables can be estimated by the user. Note that the Z_j, for $j = 1$ to p, are the result of adding correlated noise to original variables X_j. Now, by the choice of c,

$$\Sigma = \Sigma_G = \Sigma_X = (1 + \alpha)^{-1}\Sigma_Z$$

Therefore, the parameter α can be estimated as

$$\hat{\alpha} = \frac{1}{p}\sum_{j=1}^{p}(\frac{\hat{V}(\hat{Z}_j)}{\hat{V}(G_j)} - 1)$$

where $\hat{V}(\hat{Z}_j)$ and $\hat{V}(G_j)$ are the sample variances of $\hat{Z}_j$ and G_j, respectively.

Armed with the knowledge of the masked data and the distribution of the noise, which is estimated as $\sim (0, \hat{\alpha}\hat{\Sigma})$, the user can run Algorithm 1 to estimate the distribution of the original data. The rest of the attack proceeds as in Section 4.1 above.

5 Empirical Results

Consider an original data set of $n = 5000$ bivariate records $x_1, x_2, \cdots, x_{5000}$, that is, $x_i = (x_i^1, x_i^2)$ for $i = 1, \cdots, 5000$. The first variable X^1 is the annual income of the individual to which the record corresponds. The second variable X^2 is the amount of the mortgage borrowed by that individual. Figure 1 depicts the histogram of the discretized original data set: $k^1 = k^2 = 49$ are the number of intervals considered for each dimension. The histogram clearly shows that 25% of individuals (those in the lower peak of the histogram) have mortgages substantially higher and incomes substantially lower than the remaining 75% of the individuals.

Figure 2 shows the histogram of a masked data set $z_i = (z_i^1, z_i^2)$, for $i = 1, \cdots, 5000$, obtained after adding correlated noise with $\alpha = 1.167$ to the original data set. It can be seen that both peaks in the original data set collapse into a single peak in the masked data set, so that the minority with high mortgages and low income is no longer visible. Thus, publication of the masked data set protects the privacy of the individuals in that minority.

The histogram in Figure 3 has been reconstructed using Algorithm 1 on the published masked data set. The minority with low income and high mortgage is

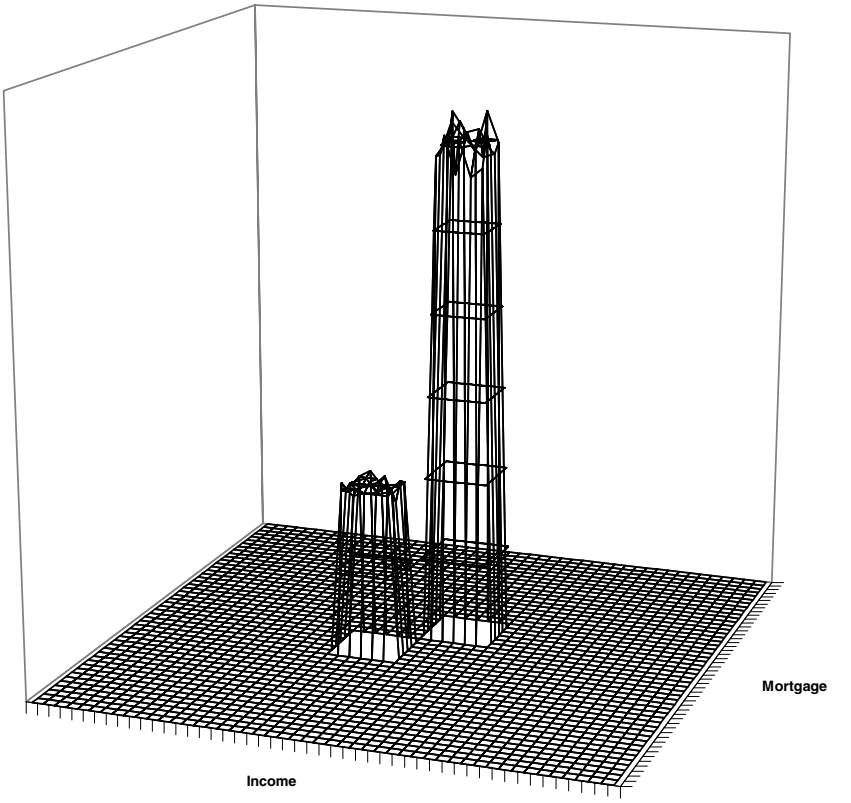

Fig. 1. Histogram of the original data set

clearly detected by the reconstruction algorithm, as can be seen from the figure. Thus, individuals in this minority are more identifiable that it would seem when merely looking at the published data set.

It may be argued that discovering that there is a 25% minority with low income and high mortgage in a population of 5000 individuals does not lead to disclosure of any particular individual. This may be true but, if more variables are available on the individuals in the data set, a reconstruction with higher p is feasible, which would lead to a segmentation of that minority and eventually to very small minorities (with a few or even a single individual). Thus, the higher the dimensionality p of the reconstruction, the more likely is disclosure.

6 Conclusions and Future Research

We have shown that, for noise addition methods used in practice, it is possible for a user of the masked data to estimate the distribution of the original data. This is so because the masking parameters must be published in order for estimates obtained from the masked data to be adjusted for consistency and unbiasedness.

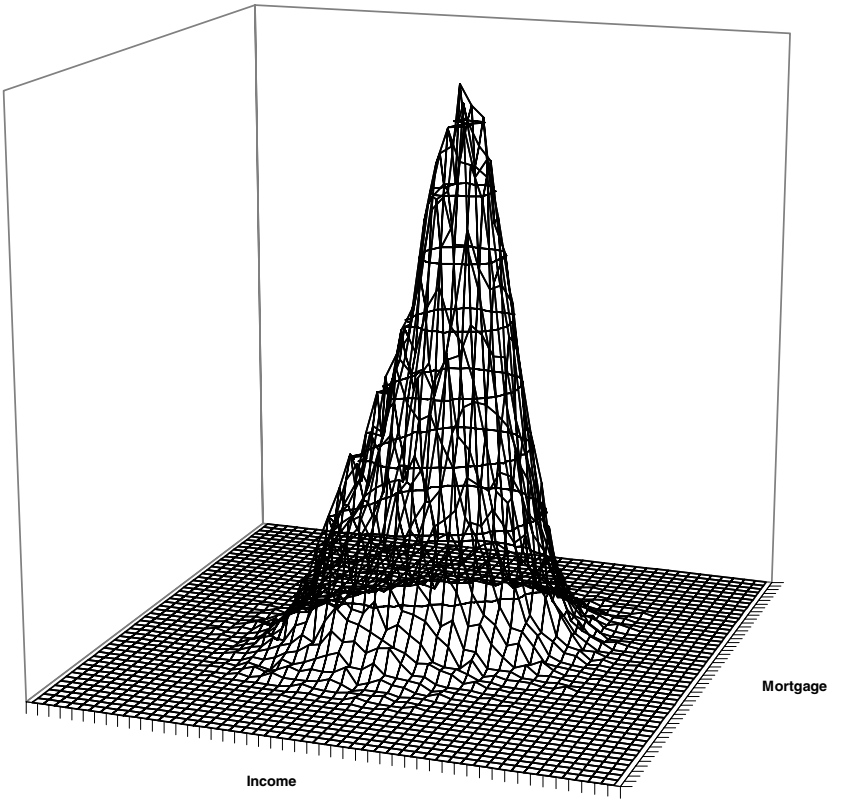

Fig. 2. Histogram of the data set masked by correlated noise addition

Estimation of the distribution of original data can lead to disclosure, as shown in the section on empirical results.

The main consequence of this work is that, as currently used, masking by noise addition may fail to adequately protect the privacy of individuals behind a microdata set.

Future research will be directed to carrying out experiments with real data sets and higher dimensionality p. A higher p is needed to increase the likelihood of partial/total reconstruction of individual original records. Since the computational requirements increase with p, substantial computing resources will be necessary for this empirical work.

Acknowledgments

This work was partly supported by the European Commission under project "CASC" (IST-2000-25069) and by the Spanish Ministry of Science and Technology and the FEDER fund under project "STREAMOBILE" (TIC-2001-0633-C03-01).

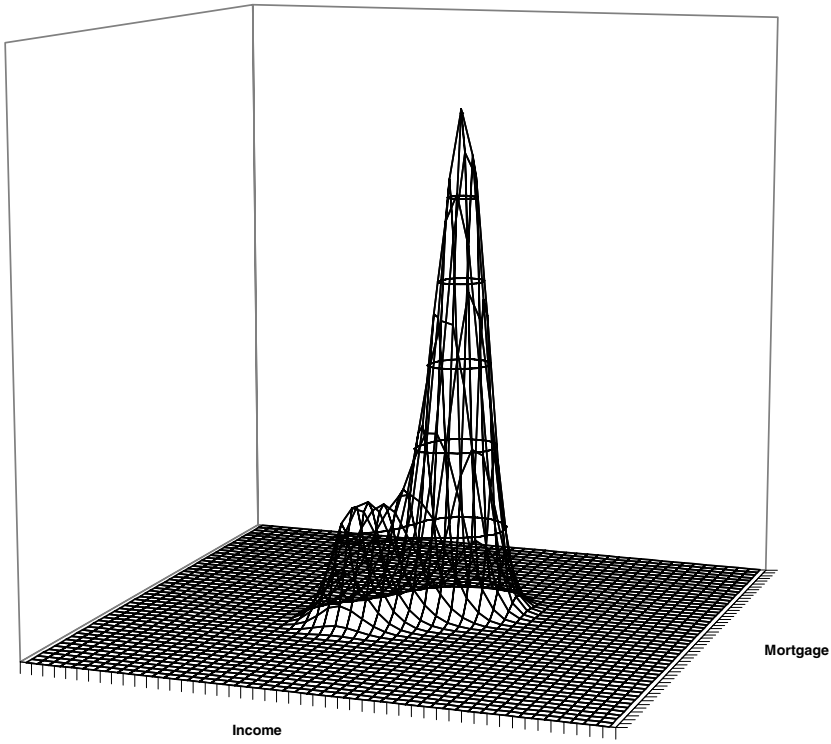

Fig. 3. Histogram of the reconstructed data set

References

1. R. Agrawal and R. Srikant, "Privacy preserving data mining", in *Proceedings of the ACM SIGMOD*, pp. 439-450, 2000.
2. D. Agrawal and C. C. Aggarwal, "On the design and quantification of privacy preserving data mining algorithms", in *Proceedings of the 20th Symposium on Principles of Database Systems*, Santa Barbara, California, USA, May 2001.
3. R. Brand, "Microdata protection through noise addition", *Lecture Notes in Computer Science*, vol. 2316, pp. 97-116, 2002. Volume title *Inference Control in Statistical Databases*, ed. J. Domingo-Ferrer. Berlin: Springer-Verlag, 2002.
4. R. Brand, "Tests of the applicability of Sullivan's algorithm to synthetic data and real business data in official statistics", Deliverable 1.1-D1, project IST-2000-25069 CASC, 28 August 2002. http://neon.vb.cbs.nl/casc
5. J. Domingo-Ferrer and V. Torra, "Disclosure protection methods and information loss for microdata", in *Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies*, eds. P. Doyle, J. Lane, J. Theeuwes and L. Zayatz. Amsterdam: North-Holland, 2001, pp. 91-110.
6. European Commission, "Open consultation on "Trust Barriers for B2B e-Marketplaces". Presentation of the main results". Brussels: EC, June 2002. Available from http://www.privacyexchange.org/iss/surveys

7. Jupiter Media Metrix, "Seventy percent of US consumers worry about online privacy, but few take protective action" (survey summary). New York: JMM, June 3, 2002. Available from http://www.jupiterresearch.com/xp/jmm/press/2002/pr_060302.html
8. H. Kargupta, S. Datta, Q. Wang and K. Sivakumar, "On the privacy preserving properties of random data perturbation techniques", in *3rd IEEE International Conference on Data Mining*, Nov. 2003, pp. 99-106. Winner of the Best Paper Award. Extended version entitled "Random data perturbation techniques and privacy preserving data mining".
9. J. J. Kim, "A method for limiting disclosure in microdata based on random noise and transformation", in *Proceedings of the Section on Survey Research Methods*. American Statistical Association, 1986, pp. 303-308.
10. J. J. Kim, "Subpopulation estimation for the masked data", in *Proceedings of the Section on Survey Research Methods*. American Statistical Association, 1990, pp. 456-461.
11. Princeton Survey Research Associates, "A matter of trust: What users want from Web sites". Yonkers, NY: Consumer WebWatch, April 16, 2002. Available from http://www.consumerwebwatch.org/news/1_abstract.htm
12. G. M. Roque, *Masking Microdata Files with Mixtures of Multivariate Normal Distributions*, unpublished Ph. D. Thesis, University of California at Riverside, 2000.
13. G. R. Sullivan, *The Use of Added Error to Avoid Disclosure in Microdata Releases*, unpublished Ph. D. Thesis, Iowa State University, 1989.
14. P. Tendick, "Optimal noise addition for preserving confidentiality in multivariate data", *Journal of Statistical Planning and Inference*, vol. 27, pp. 341-353, 1991.
15. P. Tendick and N. Matloff, "A modified random perturbation method for database security", *ACM Transactions on Database Systems*, vol. 19, pp. 47-63.
16. L. Willenborg and T. de Waal, *Elements of Statistical Disclosure Control*. New York: Springer-Verlag, 2001.

Appendix. Multivariate Generalization of the Agrawal-Srikant Reconstruction

The notation introduced in Section 3 will be used in this Appendix. Let the value of $X_i + Y_i$ be $x_i + y_i = z_i = (z_i^1, \cdots, z_i^p)$. The sum z_i is assumed to be known, but x_i and y_i are unknown. If knowledge of the densities f_X and f_Y is assumed, the Bayes' rule can be used to estimate the posterior distribution function F'_{X_1} for X_1 given that $X_1 + Y_1 = z_1 = (z_1^1, \cdots, z_1^p)$:

$$F'_{X_1}(a^1, \cdots, a^p) =$$

$$= \int_{-\infty}^{a^1} \cdots \int_{-\infty}^{a^p} f_{X_1}(w^1, \cdots, w^p | X_1 + Y_1 = (z_1^1, \cdots, z_1^p)) dw^1 \cdots dw^p$$

$$= \int_{-\infty}^{a^1} \cdots \int_{-\infty}^{a^p} \frac{f_{X_1+Y_1}(z_1^1, \cdots, z_1^p | X_1 = (w^1, \cdots, w^p)) f_{X_1}(w^1, \cdots, w^p)}{f_{X_1+Y_1}(z_1^1, \cdots, z_1^p)} dw^1 \cdots dw^p$$

$$= \frac{\int_{-\infty}^{a^1} \cdots \int_{-\infty}^{a^p} f_{X_1+Y_1}(z_1^1, \cdots, z_1^p | X_1 = (w^1, \cdots, w^p)) f_{X_1}(w^1, \cdots, w^p) dw^1 \cdots dw^p}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1+Y_1}(z_1^1, \cdots, z_1^p | X_1 = (w^1, \cdots, w^p)) f_{X_1}(w^1, \cdots, w^p) dw^1 \cdots dw^p}$$

$$= \frac{\int_{-\infty}^{a^1} \cdots \int_{-\infty}^{a^p} f_{Y_1}(z_1^1 - w^1, \cdots, z_1^p - w^p) f_{X_1}(w^1, \cdots, w^p) dw^1 \cdots dw^p}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{Y_1}(z_1^1 - w^1, \cdots, z_1^p - w^p) f_{X_1}(w^1, \cdots, w^p) dw^1 \cdots dw^p}$$

Since $f_{X_1} \equiv f_X$ and $f_{Y_1} \equiv f_Y$, the posterior distribution function F'_X given $x_1 + y_1, \cdots, x_n + y_n$ can be estimated as

$$F'_X(a^1, \cdots, a^p) = \frac{1}{n} \sum_{i=1}^{n} F'_{X_i}(a^1, \cdots, a^p)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{\int_{-\infty}^{a^1} \cdots \int_{-\infty}^{a^p} f_Y(z_i^1 - w^1, \cdots, z_i^p - w^p) f_X(w^1, \cdots, w^p) dw^1 \cdots dw^p}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_Y(z_i^1 - w^1, \cdots, z_i^p - w^p) f_X(w^1, \cdots, w^p) dw^1 \cdots dw^p}$$

Now, the posterior density function f'_X can be obtained by differentiating F'_X:

$$f'_X(a^1, \cdots, a^p) = \tag{3}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{f_Y(z_i^1 - a^1, \cdots, z_i^p - a^p) f_X(a^1, \cdots, a^p)}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_Y(z_i^1 - w^1, \cdots, z_i^p - w^p) f_X(w^1, \cdots, w^p) dw^1 \cdots dw^p}$$

Equation (3) can be used to iteratively approximate the unknown real density f_X, assuming that f_Y is known. We can take the p-dimensional uniform distribution as the initial estimate f_X^0 of f_X on the right-hand side of Equation (3); then we get a better estimate f'_X on the left-hand side. In the next iteration, we use the latter estimate on the right-hand side, and so on.

By applying the interval approximations discussed in Section 3 to Equation (3), we obtain

$$f'_X(a^1, \cdots, a^p) = \tag{4}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{f_Y(m(I_{z_i^1, \cdots, z_i^p}) - m(I_{a^1, \cdots, a^p})) f_X(I_{a^1, \cdots, a^p})}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_Y(m(I_{z_i^1, \cdots, z_i^p}) - m(I_{w^1, \cdots, w^p})) f_X(I_{w^1, \cdots, w^p}) dw^1 \cdots dw^p}$$

Let $V_{q_1, \cdots, q_p}$ be the p-dimensional volume of interval $I_{q_1, \cdots, q_p}$. We can replace the integral in the denominator of Equation (4) with a sum, since $m(I_{w^1, \cdots, w^p})$ and $f(I(w^1, \cdots, w^p))$ do not change within an interval:

$$f'_X(a^1, \cdots, a^p) =$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{f_Y(m(I_{z_i^1, \cdots, z_i^p}) - m(I_{a^1, \cdots, a^p})) f_X(I_{a^1, \cdots, a^p})}{\sum_{t^1=1}^{k^1} \cdots \sum_{t^p=1}^{k^p} f_Y(m(I_{z_i^1, \cdots, z_i^p}) - m(I_{t^1, \cdots, t^p})) f_X(I_{t^1, \cdots, t^p}) V_{t^1, \cdots, t^p}}$$

The average value of f'_X can now be computed over the interval $I_{a^1, \cdots, a^p}$ as

$$f'_X(I_{a^1, \cdots, a^p}) = \int_{I_{a^1, \cdots, a^p}} f'_X(w^1, \cdots, w^p) dw^1 \cdots dw^p / V_{a^1, \cdots, a^p} =$$

$$\int_{I_{a^1,\cdots,a^p}} \frac{1}{n}\sum_{i=1}^{n} \frac{f_Y(m(I_{z_i^1,\cdots,z_i^p}) - m(I_{w^1,\cdots,w^p}))f_X(I_{w^1,\cdots,w^p}))dw^1\cdots dw^p}{\sum_{t^1=1}^{k^1}\cdots\sum_{t^p=1}^{k^p} f_Y(m(I_{z_i^1,\cdots,z_i^t}) - m(I_{t^1,\cdots,t^p}))f_X(I_{t^1,\cdots,t^p})V_{t^1,\cdots,t^p}}$$

$$/V_{a^1,\cdots,a^p}$$

Since the above formula does not distinguish between points within the same interval, we can accumulate those points using the counters $N(\cdot)$ mentioned in Section 3:

$$f'_X(I_{a^1,\cdots,a^p}) = \tag{5}$$

$$= \frac{1}{n}\sum_{s^1=1}^{k^1}\cdots\sum_{s^p=1}^{k^p} N(I_{s^1,\cdots,s^p})\times$$

$$\times\frac{f_Y(m(I_{s^1,\cdots,s^p}) - m(I_{a^1,\cdots,a^p}))f_X(I_{a^1,\cdots,a^p})}{\sum_{t^1=1}^{k^1}\cdots\sum_{t^p=1}^{k^p} f_Y(m(I_{s^1,\cdots,s^p}) - m(I_{t^1,\cdots,t^p}))V_{t^1,\cdots,t^p}}$$

Finally, let $Pr'(X \in I_{a^1,\cdots,a^p})$ be the posterior probability of X belonging to interval $I_{a^1,\cdots,a^p}$, *i.e.*

$$Pr'(X \in I_{a^1,\cdots,a^p}) = f'_X(I_{a^1,\cdots,a^p}) \times V_{a^1,\cdots,a^p}$$

Multiplying both sides of Equation (5) by $V_{a^1,\cdots,a^p}$ and using that

$$Pr(X \in I_{a^1,\cdots,a^p}) = f_X(I_{a^1,\cdots,a^p}) \times V_{a^1,\cdots,a^p}$$

we have

$$Pr'(X \in I_{a^1,\cdots,a^p}) = \tag{6}$$

$$= \frac{1}{n}\sum_{s^1=1}^{k^1}\cdots\sum_{s^p=1}^{k^p} N(I_{s^1,\cdots,s^p})\times$$

$$\times\frac{f_Y(m(I_{s^1,\cdots,s^p}) - m(I_{a^1,\cdots,a^p}))Pr(X \in I_{a^1,\cdots,a^p})}{\sum_{t^1=1}^{k^1}\cdots\sum_{t^p=1}^{k^p} f_Y(m(I_{s^1,\cdots,s^p}) - m(I_{t^1,\cdots,t^p}))Pr(X \in I_{t^1,\cdots,t^p})}$$

Equation (6) is the basis for the recurrence used in Algorithm 1.