First extensive testing of noise addition was due to Spruill (Spruill, 1983). (Brand, 2002) gives an overview of these approaches for noise addition as well as more sophisticated tech- niques. (Domingo-Ferrer *et al.*, 2004) also describes some of the existing methods as well as the difficulties for its application in privacy. In addition to that, there exists a related approach known as multiplicative noise (see e.g. (Kim and Winkler, 2003, Liu *et al*., 2006) for details).

PRAM

PRAM, Post-RAndomization Method (Gouweleeuw *et al*., 1998), is a method for categorical data where categories are replaced according to a given probability.

Formally, it is based on a Markov matrix on the set of categories. Let $C = \{c_1, \ldots, c_c\}$ be the set of categories, then P is the Markov matrix on C when $P : C \times$

$C \rightarrow [0, 1]$ such that $\sum_{c_j \in C} P(c_i, c_j) = 1$. Then, X' is constructed from X replacing, with probability $P(c_i, c_j)$, each c_i in X by a c_j.

The application of PRAM requires an adequate definition of the probabilities $P(c_i, c_j)$. (Gouweleeuw *et al.*, 1998) proposes the Invariant PRAM. Given $T = (T(c_1) \ldots T(c_c))$ the vector of frequencies of categories in C, it consists of defining P such that frequencies are kept after PRAM. That is, $\sum_{i=1}^{c} T(c_i) p_{ij} = T(c_j)$ for all j. Then, assuming without loss of generality $T(c_k) \geq T(c_i)$ for all i, and given a parameter θ such that $0 < \theta < 1$, p_{ij} is defined as follows:

$$p_{ij} = \begin{cases} 1 - (\theta T(c_k)/T(c_i)) & \text{if } i = j \\ \theta T(c_k)/((k-1)T(c_i)) & \text{if } i \neq j \end{cases}$$

Note that a θ equal to zero implies no perturbation, and θ equal to 1 implies total perturbation. So, θ permits the user to control the degree of distortion suffered by the data set.

(Gross *et al.*, 2004) proposes the computation of matrix P from a preference matrix $W = \{w_{ij}\}$ where w_{ij} is our degree of preference about replacing category c_i by category c_j. Formally, given W the probabilities P are determined from the following optimization function:

Minimize $\sum_{i,j} w_{ij} p_{ij}$
Subject to
$p_{ij} \geq 0$
$\sum_j p_{ij} = 1$
$\sum_{i=1}^{c} T(c_i) p_{ij} = T(c_j)$ for all j

(Gross *et al.*, 2004) use integers to express preferences, and $w_{ij} = 1$ is the most preferred change, $w_{ij} = 2$ is the second most preferred changes, and so on.

Lossy Compression

This approach, first proposed in (Domingo-Ferrer and Torra, 2001a), consists of viewing a numerical data file as a grey-level image. Rows are records and columns are attributes. Then, a lossy compression method is applied to the *image*, obtaining a *compressed image*. This *image* is then decompressed and the *decompressed image* corresponds to the masked file.

Different compression rates lead to files with different degrees of distortion. I.e., the more compression, the more distortion. (Domingo-Ferrer and Torra, 2001a) used JPEG, which is based on DCT, for the compression. (Jimenez and Torra, 2009) uses JPEG 2000, which is based on wavelets.