引例:

对于属性$(bool_1, bool_2)$，存在四种可能取值，即

$$(0,0),\quad (0,1),\quad (1,0),\quad (1,1)$$

在源数据库中，四种取值的概率分别为 a,b,c,d。设有 q 概率对真实数据保持不变，$1-q$ 的概率取反（$\overline{bool_1}, \overline{bool_2}$）。操作完成后四种取值的概率分别为 t_1,t_2,t_3,t_4，则有等式

$$\begin{cases} aq + c(1-q) = c + (a-c)q = t_1 \\ bq + d(1-q) = d + (b-d)q = t_2 \\ cq + a(1-q) = a + (c-a)q = t_3 \\ dq + b(1-q) = b + (d-b)q = t_4 \\ a + b + c + d = 1 \end{cases}$$

在已知t_1,t_2,t_3,t_4 以及 q 时，上述方程组即关于a,b,c,d的四元一次方程组，有解

推广:

对于形如$(bool_1, bool_2, \ldots, bool_n)$的属性，存在 2^n 种取值，在源数据库中 2^n 种取值的概率分别为 $\{t_i\}, i = 1,2,\ldots,2^n$。设有 q 概率对真实数据保持不变，$1-q$的概率取反$(\overline{bool_1}, \overline{bool_2}, \ldots, \overline{bool_n})$。操作完成后 2^n 种取值的概率分别为 $\{p_i\}, i = 1,2,\ldots,2^n$，则有等式

$$\begin{cases} t_1 q + t_{2^n}(1-q) = t_{2^n} + (t_1 - t_{2^n})q = p_1 \\ t_2 q + t_{2^n-1}(1-q) = t_{2^n-1} + (t_2 - t_{2^n-1})q = p_2 \\ \cdots \\ t_i q + t_{2^n+1-i}(1-q) = t_{2^n+1-i} + (t_i - t_{2^n+1-i})q = p_i \\ \cdots \\ t_{2^n} q + t_1(1-q) = t_1 + (t_{2^n} - t_1)q = p_n \\ \sum_{i=1}^{2^n} t_i = 1 \end{cases}$$

在已知 $\{p_i\}$ 以及的情况下 q 时，上述方程组即关于 $\{t_i\}$ 的 2^n 元一次方程组，显然有解

在求得解为 $\{t_i\}, i = 1,2,\ldots,2^n$ 之后，考虑以下问题:

原数据分布概率为所求得解，但经处理后的数据分布概率仍为$\{p_i\}, i = 1,2,\ldots,2^n$，与原数据分布有出入，为了保证数据分布，则再次进行变换，还原回数据的原概率分布，仍然作随机替换，对于每种取值，设有 q_i 概率对真实数据保持不变，$1-q_i$ 的概率取反，则有等式

$$p_i q_i + p_{2^n+1-i}(1-q_i) = p_{2^n+1-i} + (p_i - p_{2^n+1-i})q_i = t_i$$

有解

$$q_i = \frac{t_i - p_{2^n+1-i}}{p_i - p_{2^n+1-i}}$$

此时保证了数据的分布，且由于随机替换算法，得到数据的人无法确定原数据，根据差分隐私概念，隐私得到保护。