随机响应机制效用优化研究

周异辉[1]，鲁来凤[2,3]，吴振强[1,3]

（1. 陕西师范大学计算机科学学院，陕西 西安 710119；2. 陕西师范大学数学与信息科学学院，陕西 西安 710119；
3. 贵州大学贵州省公共大数据重点实验室，贵州 贵阳 550025）

摘 要：针对本地化差分隐私中的隐私-效用均衡问题，对差分隐私和近似差分隐私情形下的二元广义随机响应机制建立效用优化模型，并采用图解法、最优性证明、软件求解和极值点等方法求解，得到了效用最优值与隐私预算、输入数据分布的精确表达式，给出了相应的效用最优机制。研究结果表明效用最优值和效用最优机制均与隐私预算和输入数据分布相关。另外，多元随机响应机制效用优化模型可通过本地化差分隐私极值点来求解。

关键词：本地化差分隐私；随机响应；效用优化；极值点；单纯形法

中图分类号：TP309

文献标识码：A

doi: 10.11959/j.issn.1000−436x.2019088

Study on utility optimization for randomized response mechanism

ZHOU Yihui[1], LU Laifeng[2,3], WU Zhenqiang[1,3]

1. School of Computer Science, Shaanxi Normal University, Xi'an 710119, China
2. School of Mathematics and Information Science, Shaanxi Normal University, Xi'an 710119, China
3. Guizhou Provincial Key Lab of Public Big Data, Guizhou University, Guiyang 550025, China

Abstract: For the study of privacy-utility trade-off in local differential privacy, the utility optimization models of binary generalized random response mechanism for the case of differential privacy and approximate differential privacy were established. By graphic method, optimality proof, software solution and extreme point method, the exact expression of the optimal utility with privacy budget and the distribution of input data was obtained, and the corresponding optimal randomized response mechanism was given. The results show that both the optimal utility and optimal mechanism are related to privacy budget and input data distribution. Moreover, the discussion for multivariate randomized response mechanism shows that the method of extreme points of local differential privacy is feasible to the solution.

Key words: local differential privacy, randomized response, utility optimization, extreme point, simplex method

1 引言

在大数据时代，用户隐私和数据安全已经成为人们普遍关注的热点问题。在各种隐私保护模型中，建立在严格数学理论基础上的差分隐私模型[1]能够量化随机机制对用户数据的隐私保护强度，保证统计数据库的查询结果不会受到任何单一用户数据的影响，因此成为当前隐私保护研究领域中备受关注的隐私保护模型之一。

差分隐私主要分为中心化差分隐私和本地化

收稿日期：2018–10–09；**修回日期**：2019–02–13

通信作者：鲁来凤，lulaifeng@snnu.edu.cn

基金项目：国家自然科学基金资助项目（No.61673251）；陕西省自然科学基金资助项目（No.2018JM6050, No.2017JQ6038）；贵州省公共大数据重点实验室开放课题基金资助项目（No.2017BDKFJJ026, No.2018BDKFJJ004）；中央高校基本科研业务费专项基金资助项目（No.GK201903091, No.GK201903011）

Foundation Items: The National Natural Science Foundation of China (No.61673251), The Natural Science Foundation of Shaanxi Province (No.2018JM6050, No.2017JQ6038), The Open Project Fund of Guizhou Provincial Key Laboratory of Public Big Data (No.2017BDKFJJ026, No.2018BDKFJJ004), The Fundamental Research Funds for the Central Universities (No.GK201903091, No.GK201903011)

差分隐私，其中中心化差分隐私是数据拥有者将数据提供给数据收集者。数据发布分为交互式和非交互式 2 种。在交互式环境下，数据分析者向数据管理者提出查询请求，数据管理者根据查询请求对数据集进行操作，并将结果进行扰动后反馈给数据分析者，数据分析者不能看到数据集的全貌，从而保护数据集中的个体隐私。在非交互式环境下，数据管理者针对所有可能的查询，在满足差分隐私的条件下一次性发布所有查询的结果；或者数据管理者发布一个不精确的数据集，即原始数据集的“净化”版本，数据分析者可以对该版本的数据集自行进行所需的查询操作。由于中心化差分隐私存在数据分析者不可信的安全隐患，因此，近年来，本地化差分隐私备受关注。在本地化差分隐私中，数据拥有者将数据进行扰动后发给数据分析者，以抵御不可信数据收集者的隐私攻击。为保证用户信息的隐私性，苹果公司于 2016 年 6 月宣布使用本地化差分隐私方法收集用户数据[2]，Google 也利用本地化差分隐私技术收集用户的行为统计数据[3]。随机响应机制是 Warner[4]于 1965 年提出的，现为本地化差分隐私保护技术的主要扰动机制，其主要思想是利用敏感信息的不确定性来保护数据信息，因此，本文针对随机响应机制的效用优化展开研究。

数据的隐私性和效用性是隐私保护技术中最重要的2个衡量指标,如何在隐私预算确定的条件下,寻找效用最高的差分隐私保护机制是许多学者关注的问题。不同研究背景采用不用的效用测度，Comas 等[5]以噪声分布在 0 附近的集中程度作为评价标准；Geng 等[6-8]以期望损失为效用测度，证明了阶梯型分布噪声是最优的；Holohan 等[9]使用输入数据分布估计误差作为效用函数，探讨了在隐私强度保证的前提下二元随机响应机制的效用优化问题。本文主要研究的问题是在确保隐私预算的前提下，分别针对 ε-差分隐私（ε-DP）和 (ε,δ)-差分隐私（(ε,δ)-DP）这 2 种隐私保护模型，研究二元随机响应机制的效用优化问题，并将其推广到多元随机响应机制。

下面举例说明本文要研究的问题。假如某校的教务处对全校师生进行问卷调查以了解其对教务管理系统的满意情况，问卷问题为“您对我校的教务管理系统是否满意”。为了保护师生的隐私，师生不必如实回答问卷调查，而是采用随机方法（比如投硬币）以某概率如实回答问卷。本文研究的问题是在保证差分隐私的前提下，以多大的概率如实回答问卷问题才能使如实回答问题的师生比例的数学期望达到最大。本文讨论的问题可抽象如下：设数据提供者的数据 x 来自输入字母表 $\overline{X}$，经扰动后输出数据为 y，属于输出字母表 $\overline{Y}$，这里只讨论 $\overline{Y}=\overline{X}$ 的情形。假设利用输入与输出相同的记录占总记录比例的数学期望作为效用测度，给定隐私预算 ε（和参量 δ），在所有满足 ε-DP（或 (ε,δ)-DP）的机制中，寻求影响效用最优机制和效用最优值的相关因素，探讨效用最优机制的条件概率矩阵和最优效用值。

2　ε-DP 下二元随机响应机制效用优化

二元随机响应机制的输入字母表 $\overline{X}=\{0,1\}$，机制可用设计矩阵（或称条件概率矩阵）$\boldsymbol{P}=\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$ 表示，其中，$p_{ij}(i,j\in\{0,1\})$ 表示输入数据为 i 时输出数据为 j 的条件概率，因此设计矩阵的行和为 1，故有

$$\boldsymbol{P}=\begin{pmatrix} p_{00} & 1-p_{00} \\ 1-p_{11} & p_{11} \end{pmatrix} \tag{1}$$

若 $\forall j\in\overline{Y}$，$\forall i_1,i_2\in\overline{X}$，有 $p_{i_1 j}\leqslant \mathrm{e}^{\varepsilon}p_{i_2 j}+\delta$，则称机制 $\boldsymbol{P}$ 满足 (ε,δ)-DP。特别地，若 δ=0，则称 $\boldsymbol{P}$ 满足 ε-DP 。

2.1　ε-DP 下的效用优化模型 I

设 π_i 表示数据 $i(i\in\{0,1\})$ 在输入数据库中的比例，则 $0\leqslant\pi_i\leqslant 1$ 且 $\pi_0+\pi_1=1$。对于离散数据，只有输出与输入相同时，数据才有价值，因此结合输入数据库的分布情况，本文用输出关于输入数据库正确率的数学期望作为效用度量，即 $u=\pi_0 p_{00}+\pi_1 p_{11}$。

优化模型 I 表示为

$$\max u=\pi_0 p_{00}+\pi_1 p_{11} \tag{2}$$

约束条件为

$$\begin{cases} 1-p_{11}\leqslant \mathrm{e}^{\varepsilon}p_{00} \\ 1-p_{00}\leqslant \mathrm{e}^{\varepsilon}p_{11} \\ p_{00}\leqslant \mathrm{e}^{\varepsilon}(1-p_{11}) \\ p_{11}\leqslant \mathrm{e}^{\varepsilon}(1-p_{00}) \\ 0\leqslant p_{00},p_{11}\leqslant 1 \end{cases} \tag{3}$$

2.2　模型求解

优化模型 I 为线性规划问题。由于变量只有 2 个，首先用图解法求解，然后用最优性判定定理进行证明，最后用 Matlab 软件求解验证所得结论。

2.2.1　图解法求解

图 1 为ε=0.1时优化模型 I 的可行域，两组平行直线的斜率分别为 $-\mathrm{e}^{\varepsilon}$ 和 $-\frac{1}{\mathrm{e}^{\varepsilon}}$，顶点分别为 A(0,1)，B$(\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1},\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1})$，C(1,0) 和 D$(\frac{1}{\mathrm{e}^{\varepsilon}+1},\frac{1}{\mathrm{e}^{\varepsilon}+1})$，目标函数等值线 $u=\pi_0 p_{00}+\pi_1 p_{11}$ 的斜率为 $-\frac{\pi_0}{\pi_1}$。从图 1 可以看出 5 种情形，分别如下。

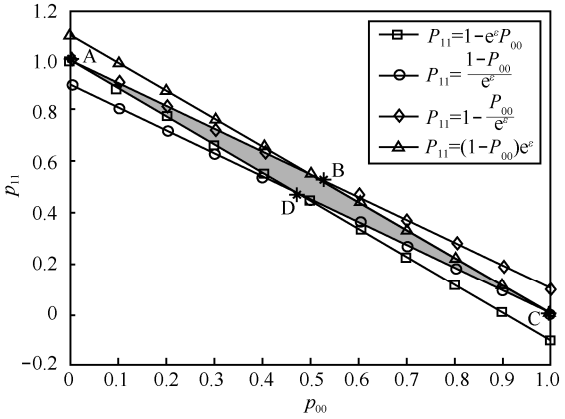


图 1　优化模型 I 的可行域(ε=0.1)

1) 当 $\frac{1}{\mathrm{e}^{\varepsilon}}\leqslant\frac{\pi_0}{\pi_1}\leqslant\mathrm{e}^{\varepsilon}$ 时，点 B 是最优解，最优值为 $\pi_0\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}+\pi_1\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}=\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}$，设计矩阵为 $\begin{pmatrix}\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1} & \frac{1}{\mathrm{e}^{\varepsilon}+1}\\ \frac{1}{\mathrm{e}^{\varepsilon}+1} & \frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}\end{pmatrix}$。

2) 当 $\frac{\pi_0}{\pi_1}\leqslant\frac{1}{\mathrm{e}^{\varepsilon}}$ 时，点 A 是最优解，最优值为 π_1=1$-\pi_0$，设计矩阵为 $\begin{pmatrix}0 & 1\\ 0 & 1\end{pmatrix}$。

3) 当 $\frac{\pi_0}{\pi_1}\geqslant\mathrm{e}^{\varepsilon}$ 时，点 C 是最优解，最优值为 π_0，设计矩阵为 $\begin{pmatrix}1 & 0\\ 1 & 0\end{pmatrix}$。

4) 特别地，当 $\frac{\pi_0}{\pi_1}=\mathrm{e}^{\varepsilon}$ 时，线段 BC 上的点都是最优解。

5) 当 $\frac{\pi_0}{\pi_1}=\frac{1}{\mathrm{e}^{\varepsilon}}$ 时，线段 AB 上的点都是最优解。

故可得最优值 u^* 与隐私预算 ε 和 π_0 的函数关系为

$$u^*=\begin{cases}\pi_0, & \pi_0\geqslant\dfrac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}\\ \dfrac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1},\dfrac{1}{\mathrm{e}^{\varepsilon}+1}\leqslant\pi_0\leqslant\dfrac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1} & \\ 1-\pi_0, & \pi_0\leqslant\dfrac{1}{\mathrm{e}^{\varepsilon}+1}\end{cases}\tag{4}$$

图 2 为按照式(4)给出的优化模型 I 的最优值 u^* 与隐私预算 ε 和输入分布 π_0 的函数关系。

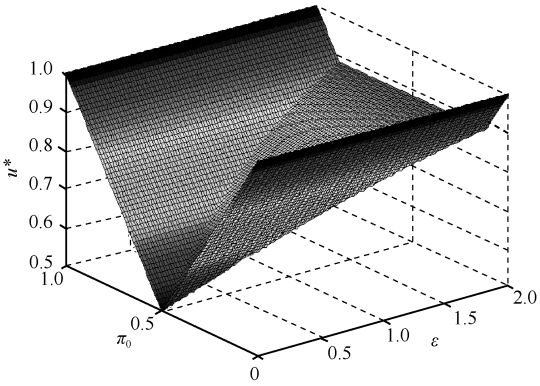


图 2　优化模型 I 中最优效用值与隐私预算和输入分布的关系

2.2.2　模型 I 最优性证明

将线性规划问题化为标准型

$$\max Z=\boldsymbol{CX}$$

约束条件为

$$\begin{cases}\boldsymbol{AX}=\boldsymbol{b}\\ \boldsymbol{X}\geqslant\boldsymbol{0}\end{cases}\tag{5}$$

其中，$\boldsymbol{A}$ 为 $m\times n$ 矩阵，秩为 m。若对于选定的基 $B=\{P_1,P_2,\cdots,P_m\}$，将式(5)所示的问题化为典则形式（简称典式）$\max Z=Z^{(0)}+\sum_{j=m+1}^{n}\sigma_j x_j$，使

$$\begin{cases}x_i+\sum_{j=m+1}^{n}a'_{ij}x_j=b'_i, i=1,2,\cdots,m\\ x_j\geqslant 0, j=m+1,m+2,\cdots,n\end{cases}\tag{6}$$

定理 1　最优性判别定理[10]。在线性规划问题的典式中，设 $\boldsymbol{X}=(b'_1,b'_2,\cdots,b'_m,0,\cdots,0)^{\mathrm{T}}$ 是对应于基 B 的一个基可行解，若有

$$\sigma_j\leqslant 0, j=m+1,m+2,\cdots,n$$

则 $\boldsymbol{X}=(b'_1,b'_2,\cdots,b'_m,0,\cdots,0)^{\mathrm{T}}$ 是优化模型 I 的最优解，并记为 $\boldsymbol{X}^*=(b'_1,b'_2,\cdots,b'_m,0,\cdots,0)^{\mathrm{T}}$，相应的目标函数最优值 $Z^*=Z^{(0)}$。

下面利用定理 1 证明式(4)确为优化模型 I 的最优值。

证明　将优化模型 I 化为标准型，其中，$\boldsymbol{X}=\left(x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8\right)^{\mathrm{T}}$，$\boldsymbol{b}=\left(1,1,\mathrm{e}^{\varepsilon},\mathrm{e}^{\varepsilon},1,1\right)^{\mathrm{T}}$，

$\boldsymbol{C}=\left(\pi_0,\pi_1,0,0,0,0,0,0\right)$，$\boldsymbol{A}=\begin{pmatrix}\mathrm{e}^{\varepsilon} & 1 & -1 & 0 & 0 & 0 & 0 & 0\\ 1 & \mathrm{e}^{\varepsilon} & 0 & -1 & 0 & 0 & 0 & 0\\ 1 & \mathrm{e}^{\varepsilon} & 0 & 0 & 1 & 0 & 0 & 0\\ \mathrm{e}^{\varepsilon} & 1 & 0 & 0 & 0 & 1 & 0 & 0\\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1\end{pmatrix}$。

为了表示的一致性，令 $x_1=p_{00}$，$x_2=p_{11}$。

根据定理 1，有

1) 取 $\boldsymbol{A}$ 的 1、2、3、4、7、8 列作为基 B_1，将优化模型 I 化为相应的典式，得到的判别系数为

$$(\sigma_5,\sigma_6)=\left(\frac{\pi_0-\pi_1\mathrm{e}^{\varepsilon}}{\mathrm{e}^{2\varepsilon}-1},\frac{\pi_1-\pi_0\mathrm{e}^{\varepsilon}}{\mathrm{e}^{2\varepsilon}-1}\right) \tag{7}$$

当 $\frac{1}{\mathrm{e}^{\varepsilon}}\leqslant\frac{\pi_0}{\pi_1}\leqslant\mathrm{e}^{\varepsilon}$ 时，判别系数全部$\leqslant 0$，对应于基 B_1 的解为基可行解，其中

$$x_1=p_{00}=x_2=p_{11}=\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}$$

此对应于图解法的第 1)情形。

2) 取矩阵 $\boldsymbol{A}$ 的 1、2、4、6、7、8 列作为基 B_2，将优化模型 I 化为相应的典式，得到的判别系数为

$$(\sigma_3,\sigma_5)=(-\frac{\pi_1-\pi_0\mathrm{e}^{\varepsilon}}{\mathrm{e}^{2\varepsilon}-1},\frac{\pi_0-\pi_1\mathrm{e}^{\varepsilon}}{\mathrm{e}^{2\varepsilon}-1}) \tag{8}$$

当 $\frac{\pi_0}{\pi_1}\leqslant\frac{1}{\mathrm{e}^{\varepsilon}}$ 时，判别系数全部$\leqslant 0$，对应于基 B_2 的解为基可行解，其中

$$x_1=p_{00}=0,\quad x_2=p_{11}=1$$

此对应于图解法的第 2)情形。

3) 取矩阵 $\boldsymbol{A}$ 的 1、2、3、5、7、8 列作为基 B_3，将优化模型 I 化为相应的典式，得到的判别系数为

$$(\sigma_4,\sigma_6)=\left(-\frac{\pi_0-\pi_1\mathrm{e}^{\varepsilon}}{\mathrm{e}^{2\varepsilon}-1},\frac{\pi_1-\pi_0\mathrm{e}^{\varepsilon}}{\mathrm{e}^{2\varepsilon}-1}\right) \tag{9}$$

当 $\frac{\pi_0}{\pi_1}\geqslant\mathrm{e}^{\varepsilon}$ 时，判别系数全部$\leqslant 0$，对应于基 B_3 的解为基可行解，其中

$$x_1=p_{00}=1,\quad x_2=p_{11}=0$$

此对应于图解法的第 3)情形。

4) 当 $\frac{\pi_0}{\pi_1}=\mathrm{e}^{\varepsilon}$ 时，$\pi_0=\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}$，$\pi_1=\frac{1}{\mathrm{e}^{\varepsilon}+1}$，线段 BC 上的点可表示为 $p_{11}=\mathrm{e}^{\varepsilon}(1-p_{00})$，其中

$$\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}\leqslant p_{00}\leqslant 1 \tag{10}$$

则线段 BC 上的点对应的效用值为

$$u=\pi_0p_{00}+\pi_1p_{11}=\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}p_{00}+\frac{1}{\mathrm{e}^{\varepsilon}+1}p_{11}=$$
$$\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}p_{00}+\frac{\mathrm{e}^{\varepsilon}(1-p_{00})}{\mathrm{e}^{\varepsilon}+1}=\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1} \tag{11}$$

此对应图解法的第 4)情形。

5) 当 $\frac{\pi_0}{\pi_1}=\frac{1}{\mathrm{e}^{\varepsilon}}$ 时，$\pi_0=\frac{1}{\mathrm{e}^{\varepsilon}+1}$，$\pi_1=\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}$，线段 AB 上的点可表示为

$$p_{11}=1-\frac{p_{00}}{\mathrm{e}^{\varepsilon}}$$

其中，$0\leqslant p_{00}\leqslant\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}$，则线段 AB 上的点对应的效用值为

$$u=\pi_0p_{00}+\pi_1p_{11}=\frac{1}{\mathrm{e}^{\varepsilon}+1}p_{00}+\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}p_{11}=$$
$$\frac{p_{00}}{\mathrm{e}^{\varepsilon}+1}+\frac{\mathrm{e}^{\varepsilon}(1-\frac{p_{00}}{\mathrm{e}^{\varepsilon}})}{\mathrm{e}^{\varepsilon}+1}=\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1} \tag{12}$$

此对应图解法的第 5)情形。

证毕。

2.2.3　软件求解验证模型 I 最优解

求解线性规划问题一般采用单纯形法，有不少现成的数学软件。利用 Matlab 中的 linprog 命令求解模型 I，并绘制最优值的图形，如图 3 所示，与图 2 对比可以发现两者是一致的。但 Matlab 软件只能对给定的隐私预算和输入分布给出相应的最优解，而不能给出函数关系的解析式。

2.3　模型Ⅰ数值仿真

图 4 为根据不同的输入数据分布情况按照式(4)给出的最优机制做出的仿真实验。比较图 2 和图 4 可以看出，仿真结果与最优值基本一致。因为效用度量是数学期望值，而仿真实验只能重复多次取平均值，所以二者不可能完全吻合。

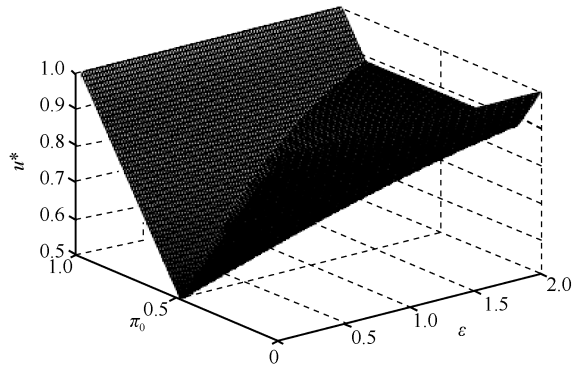


图 3 Matlab 软件求解优化模型 I 最优值的结果

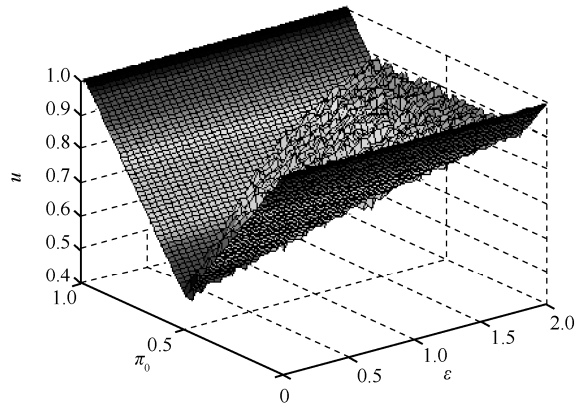


图 4 优化模型 I 效用最优值仿真结果

从式(4)和图 3 及图 4 可以看出，给定隐私预算 ε，在满足 ε-差分隐私的所有机制中，效用最优机制与输入数据库中各记录所占的比例有关：如果记录 0 所占比例超过 $\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}$，则效用最优机制为 $\begin{pmatrix}1 & 0\\ 1 & 0\end{pmatrix}$，即不管输入 0 还是 1，输出总是 0，效用最优值为 π_0；反之，如果记录 0 所占比例低于 $\frac{1}{\mathrm{e}^{\varepsilon}+1}$，则效用最优机制为 $\begin{pmatrix}0 & 1\\ 0 & 1\end{pmatrix}$，即不管输入 0 还是 1，输出总是 1，效用最优值为 $1-\pi_0$。以上 2 种情况中最优机制均为确定机制，当记录 0 所占比例介于 $\frac{1}{\mathrm{e}^{\varepsilon}+1}$ 和 $\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}$ 之间时，效用最优机制为 $\begin{pmatrix}\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1} & \frac{1}{\mathrm{e}^{\varepsilon}+1}\\ \frac{1}{\mathrm{e}^{\varepsilon}+1} & \frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}\end{pmatrix}$，即不管输入 0 还是 1，输出值与输入值相同的概率为 $\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}$，不同的概率为 $\frac{1}{\mathrm{e}^{\varepsilon}+1}$，效用最优值为 $\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}$。

3 (ε,δ)-DP 下二元随机响应机制效用优化

为避免随机机制成为确定的，可以稍微放松隐私要求，采用 (ε,δ)-DP 模型，并比较二者的效用。

3.1 (ε,δ)-DP 下的效用优化模型 II

目标函数为

$$\max u=\pi_0 p_{00}+\pi_1 p_{11} \tag{13}$$

约束条件为

$$\begin{cases}1-p_{11}\leqslant \mathrm{e}^{\varepsilon}p_{00}+\delta\\ 1-p_{00}\leqslant \mathrm{e}^{\varepsilon}p_{11}+\delta\\ p_{00}\leqslant \mathrm{e}^{\varepsilon}(1-p_{11})+\delta\\ p_{11}\leqslant \mathrm{e}^{\varepsilon}(1-p_{00})+\delta\\ 0\leqslant p_{00},p_{11}\leqslant 1\end{cases} \tag{14}$$

3.2 模型Ⅱ求解

3.2.1 图解法解模型 II

优化模型 II 可行域如图 5 所示，其中，ε=0.1，δ=0.2。顶点为 A(0,1−δ)，B(0,1)，C(δ,1)，D($\frac{\delta+\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}$, $\frac{\delta+\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}$)，E(1,$\delta$)，F(1,0)，G(1−$\delta$,0) 和 H($\frac{1-\delta}{\mathrm{e}^{\varepsilon}+1}$,$\frac{1-\delta}{\mathrm{e}^{\varepsilon}+1}$)。从 3 种情况进行分析，具体如下。

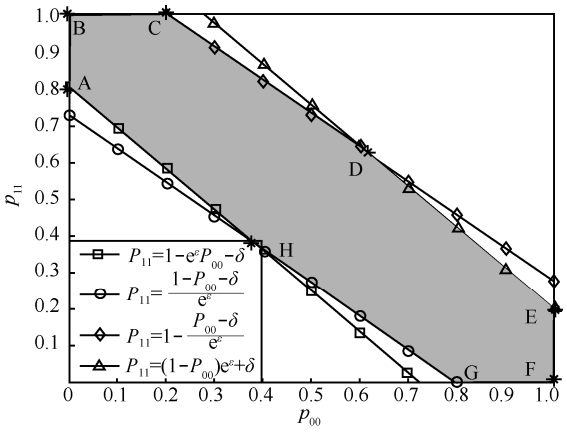


图 5 优化模型 II 的可行域（ε=0.1, δ = 0.2）

1) 当 $\frac{1}{\mathrm{e}^{\varepsilon}}\leqslant\frac{\pi_0}{\pi_1}\leqslant\mathrm{e}^{\varepsilon}$ 时，点 D 是最优解，最优值为 $\frac{\delta+\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}$，设计矩阵为 $\begin{pmatrix}\frac{\delta+\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1} & \frac{1-\delta}{\mathrm{e}^{\varepsilon}+1}\\ \frac{1-\delta}{\mathrm{e}^{\varepsilon}+1} & \frac{\delta+\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}\end{pmatrix}$。

2) 当 $\frac{\pi_0}{\pi_1}>\mathrm{e}^{\varepsilon}$ 时，点 E 是最优解，最优值为

$\pi_0+(1-\pi_0)\delta$，设计矩阵为$\begin{pmatrix}1 & 0\\ 1-\delta & \delta\end{pmatrix}$。

3) 当$\frac{\pi_0}{\pi_1}<\frac{1}{\mathrm{e}^{\varepsilon}}$时，点 C 是最优解，最优值为$\pi_0\delta+1-\pi_0$，设计矩阵为$\begin{pmatrix}\delta & 1-\delta\\ 0 & 1\end{pmatrix}$。

故最优值u^*与ε、δ和π_0的关系如式(15)所示，函数图像如图 6 所示。

$$u^*=\begin{cases}\pi_0+(1-\pi_0)\delta, \pi_0\geqslant\dfrac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}\\ \dfrac{\delta+\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}, \dfrac{1}{\mathrm{e}^{\varepsilon}+1}\leqslant\pi_0\leqslant\dfrac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}\\ \pi_0\delta+1-\pi_0,\ \pi_0\leqslant\dfrac{1}{\mathrm{e}^{\varepsilon}+1}\end{cases}\tag{15}$$

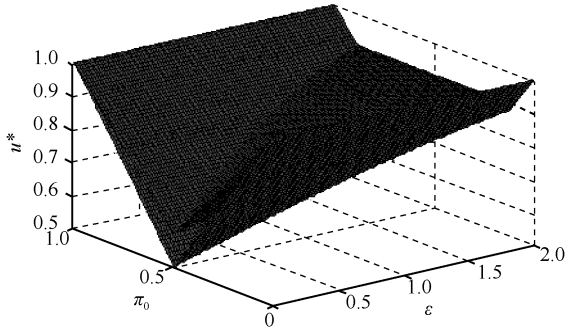


图 6　优化模型 II 中效用最优值与隐私预算和输入分布的关系

3.2.2　模型 II 最优性证明

证明　将模型 II 化为标准型，其中 $\boldsymbol{X}$、$\boldsymbol{C}$、$\boldsymbol{A}$ 与模型 I 中的一样，$\boldsymbol{b}=\left(1-\delta,1-\delta,\mathrm{e}^{\varepsilon},\mathrm{e}^{\varepsilon},1,1\right)^{\mathrm{T}}$。

1) 取 $\boldsymbol{A}$ 的 1、2、3、4、7、8 列作为基 B_1，将优化模型 II 化为相应的典式，得到的判别系数为

$$(\sigma_5,\sigma_6)=\left(\frac{\pi_0-\pi_1\mathrm{e}^{\varepsilon}}{\mathrm{e}^{2\varepsilon}-1},\frac{\pi_1-\pi_0\mathrm{e}^{\varepsilon}}{\mathrm{e}^{2\varepsilon}-1}\right)\tag{16}$$

当$\frac{1}{\mathrm{e}^{\varepsilon}}\leqslant\frac{\pi_0}{\pi_1}\leqslant\mathrm{e}^{\varepsilon}$时，判别系数全部$\leqslant 0$，对应于基 B_1 的解为基可行解，其中

$$x_1=p_{00}=x_2=p_{11}=\frac{\mathrm{e}^{\varepsilon}+\delta}{\mathrm{e}^{\varepsilon}+1}$$

此对应于图解法的第 1)情形。

2) 取矩阵 $\boldsymbol{A}$ 的 1、2、3、4、5、8 列作为基 B_2，得到的判别系数为

$$(\sigma_6,\sigma_7)=(-\pi_1,\pi_1\mathrm{e}^{\varepsilon}-\pi_0)\tag{17}$$

当$\frac{\pi_0}{\pi_1}>\mathrm{e}^{\varepsilon}$，判别系数全部$\leqslant 0$，对应于基 B_2 的解为基可行解，其中

$$x_1=p_{00}=1,\ \ x_2=p_{11}=\delta$$

此对应于图解法的第 2)种情形。

3) 取矩阵 $\boldsymbol{A}$ 的 1、2、3、4、6、7 列作为基 B_3，得到的判别系数为

$$(\sigma_5,\sigma_8)=(-\pi_0,\pi_0\mathrm{e}^{\varepsilon}-\pi_1)\tag{18}$$

当$\frac{\pi_0}{\pi_1}\leqslant\frac{1}{\mathrm{e}^{\varepsilon}}$时，判别系数全部$\leqslant 0$，对应于基 B_3 的解为基可行解，其中

$$x_1=p_{00}=\delta,\ \ x_2=p_{11}=1$$

此对应于图解法的第 3)种情形。

证毕。

3.2.3　软件求解验证模型 II 最优解

利用 Matlab 中的 linprog 命令求解优化模型 II，并绘制最优值的图形，如图 7 所示，与图 6 对比发现两者一致。

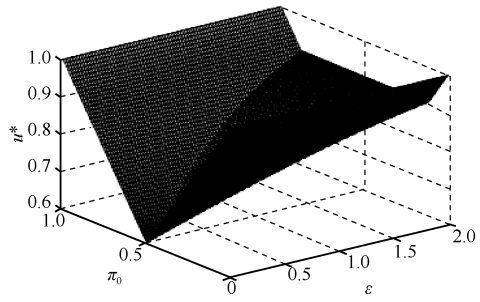


图 7　Matlab 软件求解优化模型 II 最优值结果

3.3　模型 II 数值仿真

图 8 为根据不同的输入数据分布情况按照式(15)中的最优机制做出的仿真实验。比较图 6 和图 8 可以看出，仿真结果与最优效用值一致、而且从图 8 可以看出，(ε,δ) -DP 避免了 ε-DP 中当 π_0 足够大或足够小时，机制退化为确定机制的缺点。

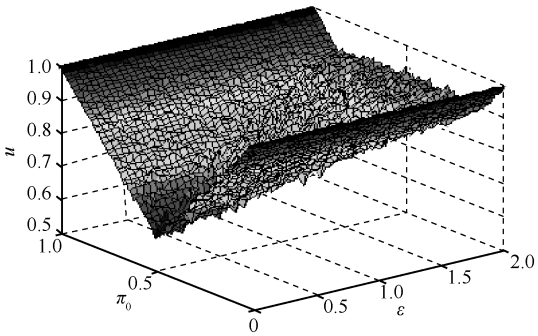


图 8　优化模型 II 最优效用值仿真结果

从式(15)和图 7 及图 8 可以看出，给定隐私预算 ε 和参量 δ，在满足 (ε,δ)-DP 的所有机制中，效

用最优机制与输入数据库中各记录所占的比例有关：如果记录 0 所占比例超过 $\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}$，则效用最优机制为 $\begin{pmatrix} 1 & 0 \\ 1-\delta & \delta \end{pmatrix}$，即输入 0 时输出总是 0，输入 1 时，输出 1 的概率为 δ，最优效用值为 $(\pi_0+(1-\pi_0)\delta)$；反之，如果记录 0 所占比例低于 $\frac{1}{\mathrm{e}^{\varepsilon}+1}$，则最优效用机制为 $\begin{pmatrix} \delta & 1-\delta \\ 0 & 1 \end{pmatrix}$，即输入 1 时输出总是 1，输入 0 时，输出是 0 的概率为 δ，效用最优值为 $\pi_0\delta+1-\pi_0$；当记录 0 所占比例介于 $\frac{1}{\mathrm{e}^{\varepsilon}+1}$ 和 $\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}$ 之间时，最优效用机制为 $\begin{pmatrix} \frac{\delta+\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1} & \frac{1-\delta}{\mathrm{e}^{\varepsilon}+1} \\ \frac{1-\delta}{\mathrm{e}^{\varepsilon}+1} & \frac{\delta+\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1} \end{pmatrix}$，即不管输入 0 还是 1，输出值与输入值相同的概率为 $\frac{\delta+\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}$，不同的概率为 $\frac{1-\delta}{\mathrm{e}^{\varepsilon}+1}$，效用最优值为 $\frac{\delta+\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}$。

对比 ε-DP 和 (ε,δ)-DP 这 2 种情形可以发现，最优机制中 (ε,δ)-DP 情形的效用比 ε-DP 情形好，但会损失一些隐私保护程度。以第 1 节的例子为例，设全校师生共 10 000 名，$\varepsilon=0.1$，δ=0.15。用“0”表示对教务系统不满意，“1”表示对教务系统满意。假设 $\pi_0=0.2$，$\pi_1=0.8$。因为 $\pi_0=0.2<\frac{1}{\mathrm{e}^{\varepsilon}+1}=0.475\,0$，所以效用最优的 ε-DP 机制的设计矩阵为 $\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$，即所有人都回答“满意”，最优效用值为 0.8。效用最优的 (ε,δ)-DP 机制的设计矩阵为 $\begin{pmatrix} 0.15 & 0.85 \\ 0 & 1 \end{pmatrix}$，最优效用值为 0.83。对其他比例的 π_0 和 π_1，也可得到相应的结论。

4 多元本地化差分隐私效用

设输入字母表 $X=\{0,1,\cdots,n-1\}$，n 元广义随机响应机制为

$$\boldsymbol{P}=\begin{pmatrix} p_{00} & p_{01} & \cdots & p_{0,n-1} \\ p_{10} & p_{11} & \cdots & p_{1,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n-1,0} & p_{n-1,2} & \cdots & p_{n-1,n-1} \end{pmatrix}$$

效用优化模型的目标函数为

$$\max u=\sum_{i=0}^{n-1}\pi_i p_{ii} \tag{19}$$

约束条件为

$$\begin{cases} p_{i_1 j}\leqslant \mathrm{e}^{\varepsilon}p_{i_2 j}+\delta, \\ \forall i_1,i_2,i_1\neq i_2,\ j\in\{0,1,\cdots,n-1\} \\ \sum\limits_{k=0}^{n-1}p_{ik}=1,\forall i\in\{0,1,\cdots,n-1\} \\ p_{ij}\geqslant 0,\forall i,j\in\{0,1,\cdots,n-1\} \end{cases} \tag{20}$$

这里有 n^2 个变量，$n^2(n-1)$ 个差分隐私限制，n 个行和为 1 限制，n^2 个非负限制。因为 $\forall i\in\{0,1,\cdots,n-1\}$，$\sum\limits_{k=0}^{n-1}p_{ik}=1$，所以每一行可以保留($n$−1)个变量，另一个变量用其他元素表示。比如，令 $p_{0,n-1}=1-\sum\limits_{k=0}^{n-2}p_{0k}$，$p_{i,i-1}=1-\sum\limits_{k=0,k\neq i-1}^{n-1}p_{ik}$，$\forall i\in\{1,\cdots,n-1\}$。这样有 $n(n-1)$ 个变量，$n^2(n-1)$ 个差分隐私限制，n 个小于或等于 1 的限制和 $n(n-1)$ 个非负限制。

首先，由于变量个数比较多，不能用图解法求解。其次，由于最优性判定定理只是最优性判定的充分条件，且与基的选择有关系，只能用于检验某自变量取值是否为最优解。随着元数的增加，自变量个数也越来越多，导致给出最优效用值与差分隐私预算及输入数据集分布的关系越来越难。再者，虽然各种数学软件求解线性规划问题很容易，但是只是针对给定的系数，而不能得到最优值与隐私预算及数据分布间关系的解析式。

然而，线性规划问题的最优解在可行域的极值点处取得，因此对差分隐私可行域极值点的研究是广大学者研究的重点[11-12]。本文采用极值点法对优化模型 I 求解。

证明 文献[11]中的结论：如果 n 元差分隐私机制 $\boldsymbol{P}$ 中存在 $p_{ij}=0$，则第 j 列元素全为 0，换句话说，差分隐私机制的列向量要么全部为 0，要么全部为非 0。文献[12]中的结论：n 元差分隐私机制恰有一个非 0 列的极值点为只有一列全部为 1 其他元素全部为 0 的矩阵；恰有 2 个非 0 列的极值点，机制中非 0 列元素为 $\frac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+1}$ 或 $\frac{1}{\mathrm{e}^{\varepsilon}+1}$。根据上述结论可得，二元本地化差分隐私可行域的所有极值点为

$\begin{pmatrix}1 & 0\\ 1 & 0\end{pmatrix}$、$\begin{pmatrix}0 & 1\\ 0 & 1\end{pmatrix}$、$\begin{pmatrix}\frac{e^{\varepsilon}}{e^{\varepsilon}+1} & \frac{1}{e^{\varepsilon}+1}\\ \frac{1}{e^{\varepsilon}+1} & \frac{e^{\varepsilon}}{e^{\varepsilon}+1}\end{pmatrix}$和$\begin{pmatrix}\frac{1}{e^{\varepsilon}+1} & \frac{e^{\varepsilon}}{e^{\varepsilon}+1}\\ \frac{e^{\varepsilon}}{e^{\varepsilon}+1} & \frac{1}{e^{\varepsilon}+1}\end{pmatrix}$，对应的效用值为$\pi_0$、$1-\pi_0$、$\frac{e^{\varepsilon}}{e^{\varepsilon}+1}$和$\frac{1}{e^{\varepsilon}+1}$。因为$\varepsilon \geqslant 0$，所以$\frac{e^{\varepsilon}}{e^{\varepsilon}+1} \geqslant \frac{1}{e^{\varepsilon}+1}$，因此只需比较$\pi_0$、$1-\pi_0$和$\frac{e^{\varepsilon}}{e^{\varepsilon}+1}$的大小，具体如下。

1) 当$\pi_0 \geqslant \frac{e^{\varepsilon}}{e^{\varepsilon}+1}$时，$1-\pi_0 \leqslant \frac{1}{e^{\varepsilon}+1} \leqslant \frac{e^{\varepsilon}}{e^{\varepsilon}+1}$，最优值为$\pi_0$。

2) 当$\pi_0 \geqslant \frac{1}{e^{\varepsilon}+1}$时，$1-\pi_0 \geqslant \frac{e^{\varepsilon}}{e^{\varepsilon}+1}$，最优值为$1-\pi_0$。

3) 当$\frac{1}{e^{\varepsilon}+1} \leqslant \pi_0 \leqslant \frac{e^{\varepsilon}}{e^{\varepsilon}+1}$时，最优值为$\frac{e^{\varepsilon}}{e^{\varepsilon}+1}$。此与式(4)一致。

证毕。

5　结束语

本文在大数据环境中隐私泄露严重、隐私保护需求日益增强的背景下，针对差分隐私中随机响应机制的效用优化问题展开研究。首先研究了广义二元随机响应机制的效用优化问题，分别针对ε-DP和(ε,δ)-DP 情形建立效用优化模型并求解，得到了最优解与隐私预算和输入数据分布的解析式，给出了相应的最优机制，并通过数值仿真验证所得结论。针对多元广义差分隐私的效用优化问题展开讨论，用差分隐私可行域的极值点去研究最优效用，其中多元随机响应机制的效用最优值与输入分布和隐私预算间的函数表达式有待进一步研究。

参考文献：

[1] DWORK C, ROTH A. Thealgorithmic foundations of differential privacy[J]. Foundations and Trends in Theoretical Computer Science, 2014, 9(3-4): 211-407.

[2] TANG J, KOROLOVAA, BAI X, et al. Privacy loss inapple's implementation of differential privacy on MacOS 10.12[J]. Cornell University, arXiv Preprint , arXiv:1709.02753, 2017.

[3] ERLINGSSON Ú, PIHUR V, KOROLOVA A. RAPPOR: randomized aggregatable privacy-preserving ordinal response[C]//The ACM SIGSAC Conference on Computer and Communications Security. ACM, 2014: 1054-1067.

[4] WARNER S L. Randomized response: a survey technique for eliminating evasive answer bias[J]. Journal of the American Statistical Association, 1965, 60(309): 63-69.

[5] COMAS J S, FERRER J D. Optimal data-independent noise for differential privacy[J]. Information Sciences, 2013, 250: 200-214.

[6] GENG Q, KAIROUZ P, OH S, et al. The staircase mechanism in differential privacy[J]. IEEE Journal of Selected Topics in Signal Processing, 2015, 9(7): 1176-1184.

[7] GENG Q, VISWANATH P. The optimal noise-adding mechanism in differential privacy[J]. IEEE Transactions on Information Theory, 2016, 62(2): 925-951.

[8] GENG Q, VISWANATH P. Optimal noise adding mechanisms for approximate differential privacy[J]. IEEE Transactions on Information Theory, 2016, 62(2): 952-969.

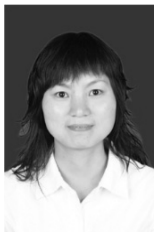
[9] HOLOHAN N, LEITH D J, MASON O. Optimal differentially private mechanisms for randomized response[J]. IEEE Transactions on Information Forensics & Security,2017, 12(11): 2726-2735.

[10] 邓成梁. 运筹学的原理和方法: 第三版[M]. 武汉: 华中科技大学出版社, 2014.
DENG C L. The Principle and Method of Operations Research [M]. 3rd ed. Wuhan: Huazhong University of Science& Technology Press, 2014.

[11] KAIROUZ P, OH S, VISWANATHP. Extremal mechanisms for local differential privacy[J]. Journal of Machine Learning Research, 2016, 4 (1): 492-542.

[12] HOLOHAN N, LEITHD J, MASON O. Extreme points of the local differential privacy polytope[J]. Linear Algebra and Its Applications, 2017, 534: 78-96.

[作者简介]



周异辉（1981− ），女，河北蠡县人，博士，陕西师范大学讲师，主要研究方向为网络安全、大数据环境下的隐私保护和差分隐私。

鲁来凤（1979− ），女，安徽桐城人，博士，陕西师范大学副教授，主要研究方向为网络安全、大数据环境下的隐私保护和差分隐私。

吴振强（1968− ），男，陕西柞水人，博士，陕西师范大学教授、博士生导师，主要研究方向为网络数据科学、纳米网络、分布式计算、隐私保护、可信计算等。