適応的反復メカニズムによる合成データ生成の評価

三浦 尭之[1,a)]　竹内 弘史[2]　宮下 朋也[2]　紀伊 真昇[1]　山﨑 雄輔[1]　芝原 俊樹[1]　石原 一郎[3]

概要：差分プライベートな合成データ生成は多くの手法が提案されており，中でも McKenna らによって提案された適応的反復メカニズム（Adaptive and Iterative Mechanism, AIM）は多くのベンチマーク論文で優れた手法として取りあげられている．しかし，AIM のどの要素が合成データの品質を向上させているのか，また，AIM の品質向上の限界については明らかではない．本稿では，AIM の内容を解説しつつ，優れた品質のデータを生成する要因分析を行う．具体的にはパラメーター初期化の工夫やノイズ付加アルゴリズムの有効性を調査した．また，zero-Concentrated 差分プライバシーによる予算管理を，よりタイトな評価が可能な数値的合成則に焼き直して評価し，同様の結果になることを明らかにした．加えて，陽に属性間の相互情報量を学習する PrivBayes よりも，AIM の方が元データの属性間の相互情報量を保存していることを実験により確かめた．

キーワード：差分プライベート合成データ生成，適応的反復メカニズム，AIM，Private-PGM, 合成則

Evaluation of Synthetic Data by Adaptive and Iterative Mechanism

Takayuki Miura[1,a)]　Hiroshi Takeuchi[2]　Tomoya Miyashita[2]　Masanobu Kii[1]
Yusuke Yamasaki[1]　Toshiki Shibahara[1]　Ichiro Ishihara[3]

Abstract: Many methods for differentially private synthetic data generation have been proposed, among which the Adaptive and Iterative Mechanism (AIM), introduced by McKenna et al., has been highlighted in numerous benchmark studies as a state-of-the-art approach. However, it remains unclear which factors of AIM drives an improvement in synthetic data quality, as well as the extent to which its quality can be further improved. In this paper, we provide an explanation of AIM and conduct a factor analysis to identify what enables it to generate high-quality data. Specifically, we investigate the impact of parameter initialization strategies and the effectiveness of its noise-adding algorithm. Furthermore, while AIM manages the privacy budget using the composition theorem of zero-concentrated differential privacy (zCDP), we demonstrate that even when this is reformulated from the perspective of numerical composition theorem, equally tight budget management is achieved. In addition, our experimental results confirm that AIM preserves mutual information between attributes in the original data more effectively than PrivBayes, which explicitly models attribute dependencies.

Keywords: Differentially Private Synthetic Data Generation, Adaptive and Iterative Mechanism, AIM, Private-PGM, Composition Theorem

1. はじめに

プライバシーリスクをコントロールしながら個人に関わる情報を公開する手法として，差分プライベートな合成データ生成（DP 合成）が提案されている [7–10]．DP 合成では，プライバシー保護のため生成パラメータなどにノイズを加える必要があり，プライバシー保護性と合成データの有用性にトレードオフがある．

McKenna らによって提案された適応的反復メカニズム（Adaptive and Iterative Mechanism, AIM [9]）がデータの有用性の観点で優れた DP 合成手法であるとして注目

[1] NTT 社会情報研究所, NTT Social Informatics Laboratories
[2] 株式会社アーク情報システム, ARK Information Systems, INC.
[3] ＮＴＴテクノクロス株式会社, NTT TechnoCross Corporation
[a)] tkyk.miura@ntt.com

されている．この手法は，後述するプライバシー損失予算を適応的に管理しながら，Private-PGM（Probabilistic Graphical-Model）というグラフィカルモデルの学習を進めることが特徴である．この手法は多くの文献で優れた合成データを生成しており [3, 13, 18]，広く用いられている PrivBayes [17] よりも優れた手法として報告されている．しかし，AIM は最適なのか，また，品質の良いデータを生成する要因などは明らかになっていない．加えて，AIM の各工夫点の有効性は提案論文の中で論じられているが，具体的な数値で定量的に評価しているわけではない．

本稿では AIM および Private-PGM の方法を解説しつつ，各工夫点の有効性を調査する．AIM の工夫点としては大きく「初期化」「ガウシアンメカニズム」「zCDP による予算管理」があげられる．それらの有効性を比較実験を行い調査した．初期化やガウシアンメカニズムについては合成データの有用性の観点からの評価を行った．初期化の工夫については，その有無が合成データの品質にあまり有効な差を与えていないという結果になった．ガウシアンメカニズムについては，ラプラスメカニズムに取り替えたところ，有用性が大幅に落ちることが確認できた．

差分プライバシーはそのプライバシー保護度合いを定量的に表現することができ，その指標をプライバシー損失予算という．また，プライバシー保護メカニズムには合成則という性質があり [6]，複数のメカニズムを組み合わせた処理全体のプライバシー損失予算を個々のメカニズムのプライバシー損失予算で表現できる．逆に言うと，全体の予算を決めた後，それ実現するような個々のメカニズムの予算を決めることもでき，その予算管理の方法がメカニズム全体の性能に大きな影響を与えることがある．AIM は zero-Concentrated 差分プライバシー（zCPD）に基づいて予算管理が行われているが，これを誤差保証付で解析的にプライバシー損失予算を計算可能な数値的合成則 [5] に焼き直して AIM の予算管理を観察した．結果として AIM の予算管理は数値的合成則と同様のタイトさでプライバシー損失予算を管理できていることが確かめられた．

最後に元データの属性間の相互情報量の傾向が合成データにも保持されているかを調査した．結果として，AIM が PrivBayes と比べて，属性間の相互情報量を保持できてることが確かめられた．PrivBayes は属性間の相互情報量に基づいて，学習するべき同時分布を決定しているが，相互情報量を陽には使わない AIM の方が相互情報量を保持している結果になった．

評価した範囲内では従来の AIM の各工夫点は少なくとも無駄にはなっておらず，デフォルトの設定で使用することが望ましいということが確かめられた．PrivBayes を使って合成データの生成がされていたケースにおいて，今後，本稿の定式化・説明を参考に AIM が使用されていくことを期待する．

2. 準備

本稿の説明に必要な記号と Private-PGM，AIM などの手法を説明する．

2.1 記法

本稿では，テーブル形式のデータセットを対象とする．行が個人に対応し，列が属性に対応すると考える．属性はカテゴリ属性，数値属性に分かれるが，数値属性は適当な離散化によりカテゴリ属性に変換可能なため，すべての属性がカテゴリ属性と仮定する．

テーブルの属性を $A_1, \ldots, A_d$ とする．レコード数は m とする．また，本稿では興味のある属性の組からなる集合を**測定対象** $\mathcal{W} \subset 2^{[d]}$ と呼ぶ．i 番目の属性の属性値の個数を $n_i \in \mathbb{Z}_{\geq 0}$ とおき，特に，$r \in \mathcal{W}$ に対して，$n_r := \prod_{i \in r} n_i$ と定める．全体は，$n := n_{[d]}$ とおく．カテゴリ属性 A_i の要素はそれぞれ適当な One-hot ベクトルに変換することで，$A_i \subset \mathbb{R}^{n_i}$ とみなせるため，データセットは $D \in \mathbb{R}^{m \times n}$ と表現できる．特に，レコードに対応するベクトルの確率分布は $p \in \mathbb{R}^n$ として持つことができる．

データセット D に対して，$r \subset [d]$ 成分の頻度分布（クロス集計）を $p_r(D) \in \mathbb{R}^{n_r}$ と表す．また，測定対象 $\mathcal{W}$ に対してその細分 $\mathcal{W}_+$ を次のように定める

$$\mathcal{W}_+ := \{r \in 2^{[d]} \mid r \subset s, s \in \mathcal{W}\}.$$

2.2 差分プライバシー

本稿で必要な差分プライバシーの定義・事実を説明する．

定義 2.1 (隣接データセット)**.** データセット $D \in \mathcal{D}$ に対して，D のレコードを 1 人分だけ変えたデータセット D' を D の**隣接データセット**と呼ぶ．D の隣接データセット全体の集合を $N(D) \subset \mathcal{D}$ とおく．

定義 2.2 (差分プライバシー [4])**.** ランダム化関数 $\mathcal{M} : \mathcal{D} \to \mathcal{Y}$ が (ε, δ)**-差分プライバシー**（(ε, δ)*-DP*）を満たすとは次が成り立つことを言う．任意の隣接データセット $D \in \mathcal{D}$，任意の $D' \in N(D)$ と任意の出力 $S \subset \mathcal{Y}$ に対して，

$$\Pr[\mathcal{M}(D) \in S] \leq e^{\varepsilon} \Pr[\mathcal{M}(D') \in S] + \delta$$

が成り立つ．ε や δ をプライバシー損失予算と呼ぶ．

定義 2.3 (zero-Concentrated 差分プライバシー [2])**.** $\rho > 0$ とする．ランダム化関数 $\mathcal{M} : \mathcal{D} \to \mathbb{R}^d$ が ρ**-zero-concentrated 差分プライバシー** *(ρ-zCDP)* を満たすとは，次が成り立つことを言う．任意の $\alpha > 1$ と任意のデータセット $D \in \mathcal{D}$，任意の $D' \in N(D)$ に対して，

$$D_\alpha(\mathcal{M}(D) || \mathcal{M}(D')) \leq \rho \cdot \alpha$$

が成り立つ．ここで，$D_\alpha(\cdot||\cdot)$ は *Renyi-divergence* である．

命題 2.4 (ρ-zCDP の (ε,δ)-DP への変換 [12])．プライバシー保護メカニズム $\mathcal{M}$ が ρ-*zCDP* を満たすなら，任意の $\varepsilon \geq 0$ について，以下の δ について (ε,δ)-*DP* を満たす；

$$\delta = \inf_{\alpha>1} \frac{e^{(\alpha-1)(\alpha\rho-\varepsilon)}}{\alpha-1}\left(1-\frac{1}{\alpha}\right)^{\alpha}.$$

定義 2.5 (ガウシアンメカニズム)．あるクエリ関数 $f:\mathcal{D}\to\mathbb{R}^d$ に対して，敏感度を

$$\Delta_2 := \sup_{D\in\mathcal{D},D'\in N(D)} \|f(D)-f(D')\|_2$$

とおく．平均が 0 で分散 σ^2 の標準正規分布 $\mathcal{N}(0,\sigma^2 I_d)$ に従うランダムノイズを f の出力に加えたランダム化関数をガウシアンメカニズムと呼ぶ．このとき，$\sigma^2 > \Delta_2^2/(2\rho)$ が成り立つなら，メカニズムは ρ-*zCDP* を満たす [2].

定義 2.6 (指数メカニズム [11])．$\mathcal{C}=\{c_1,\ldots,c_t\}$ を出力候補集合とする．スコア関数 $s:\mathcal{D}\times\mathcal{C}\to\mathbb{R}$ に対して，

$$\Delta := \sup_{D\in\mathcal{D},D'\in N(D),c\in\mathcal{C}} |s(D,c)-s(D',c)|$$

とおく．次のようなランダム化関数 $\mathcal{M}:\mathcal{D}\to\mathcal{C}$ で

$$\Pr[\mathcal{M}(D)=c] = \frac{\exp(\frac{\varepsilon}{2\Delta}s(D,c))}{\sum_{c'\in\mathcal{C}}\exp(\frac{\varepsilon}{2\Delta}s(D,c'))}$$

となるものを**指数メカニズム**と呼ぶ．このメカニズムは ε-*DP* を満たし，文献 [9] によると，$\frac{\varepsilon^2}{8}$-*zCDP* も満たす.

命題 2.7 (zCDP の合成定理)．$\mathcal{M}_1:\mathcal{D}\to\mathbb{R}^{d_1}$ が ρ_1-*zCDP*, $\mathcal{M}_2:\mathcal{D}\times\mathbb{R}^{d_1}\to\mathbb{R}^{d_2}$ が ρ_2-*zCDP* を満たすとする．このとき，次の適応的合成メカニズム $\mathcal{M}:\mathcal{D}\to\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}$

$$\mathcal{M}(D) = (\mathcal{M}_1(D), \mathcal{M}_2(D,\mathcal{M}_1(D)))$$

は $(\rho_1+\rho_2)$-*zCDP* を満たす.

定義 2.8 (プライバシー曲線)．Ω 上の 2 つの確率変数 X,Y に対して，関数 $\delta(X\|Y):\mathbb{R}\to[0,1]$ を

$$\delta(X\|Y)(\varepsilon) = \sup_{S\subseteq\Omega} \Pr[Y\in S] - e^{\varepsilon}\Pr[X\in S]$$

と定める．メカニズム $\mathcal{M}$ が (ε,δ)-*DP* を満たすことと，任意の隣接データセット D,D' に対して

$$\delta\big(\mathcal{M}(D)\,\|\,\mathcal{M}(D')\big)(\varepsilon) \leq \delta$$

が成り立つことは同値である.

メカニズム $\mathcal{M}_1,\ldots,\mathcal{M}_k$ に対して，それぞれのプライバシー曲線を $\delta_{\mathcal{M}_i}$ とおく．これらの適応的合成メカニズム $\mathcal{M}=\mathcal{M}_k\circ\cdots\circ\mathcal{M}_1$ のプライバシー曲線 $\delta_{\mathcal{M}}$ の近似計算に対する計算量を，Gopi らは導出した [5]．これは数値的合成則と呼ばれる．実装は下記などが参考にできる[*1].

[*1] https://github.com/microsoft/prv_accountant

Algorithm 1 Private-PGM

Input: 損失関数 $L(\mu)$，θ：パラメータ，(η_t)：学習スケジュール
Output: Estimated data distribution $\hat{p}_\theta$

```
for t = 1, ..., T do
    μ = MARGINAL-ORACLE(θ)
    θ = θ − η_t ∇L(μ)
return θ
```

2.3 Private-PGM

Private-PGM とは McKenna らによって提案された手法で，次で定義されるグラフィカルモデルを鏡像降下法で学習する手法である [10].

定義 2.9 (グラフィカルモデル)．$\mathcal{C}\in 2^{[d]}$ を測定対象とする．ここで，各 $c\in\mathcal{C}$ に対して，$\theta_c\in\mathbb{R}^{n_c}$ をパラメータとする $A_1\times\cdots\times A_d$ 上の確率分布

$$p_\theta(x) := \frac{1}{Z}\exp\left(\sum_{c\in\mathcal{C}}\theta_c(x_c)\right)\in\mathbb{R}$$

のことを**グラフィカルモデル**と呼ぶ．ここで, Z は正規化定数である．$x_c\in A_{r_1}\times\cdots\times A_{r_t}$ はベクトル $x\in A_1\times\cdots\times A_d$ の $c=\{r_1,\ldots,r_t\}$ 成分のみを取り出したもので, $\theta_c(x_c)\in\mathbb{R}$ はパラメータにおける x_c 成分を取り出した実数値である.

定義 2.10 (MARGINAL-ORACLE)．パラメータ $\theta=(\theta_c)_{c\in\mathcal{C}}$ を入力とし, 各 $c\in\mathcal{C}$ に対応するグラフィカルモデル $p_\theta(x)$ の周辺分布 $\mu_c\in\mathbb{R}^{n_c}$ を返す関数を **MARGINAL-ORACLE** と呼ぶ．$\mu(\theta):=(\mu_c)_{c\in\mathcal{C}}$ と書くこととする.

Private-PGM（Algorithm 1）では次のように損失関数を定める.

$$L(\mu(\theta)) := \sum_{c\in\mathcal{C}}\|\mu_c-\tilde{y}_c\|_2^2.$$

ここで，$(\tilde{y}_c)_{c\in\mathcal{C}}$ はノイズを足した測定値である．この損失関数を用いて鏡像降下法を行い，グラフィカルモデルのパラメータを更新する．すなわち，この損失関数を μ で微分して，$\nabla_\mu L(\mu)$ を得て，これによって θ を更新する．具体的な方法は A.1 節に記載した.

2.4 AIM

AIM は Private-PGM を適応的に使いながらグラフィカルモデルを学習する手法である（Algorithm 2）．AIM の入力は，データセット，測定対象，プライバシー損失予算，およびハイパーパラメータであり，その内部処理は大きく下記の通りである．(1) 仮の全体の同時分布 θ_t を初期化アルゴリズムで作る（Algorithm 3）．(2) 元データセット D と暫定分布 θ_t の差をスコアとした指数メカニズムによって，測定対象の属性 $r_t\in W$ を選ぶ．(3) 周辺分布 $p_{r_t}(D)$ にパラメータ σ_t^2 のガウシアンノイズを加える．(4) Private-PGM を用いて，パラメータ分布 θ_t を θ_{t+1} に更新．(5) プライバシー予算をアニーリング（Algorithm 4）．更新量が小さいとき予算を使うペースを大きくして，更新量

Algorithm 2 AIM：適応的反復メカニズム

Input: D：データセット, $\mathcal{W}$：測定対象, ρ：プライバシー損失予算, 配分ペース：$T = 16d$，配分割合：$\alpha = 0.9$
Output: $\hat{D}$：合成データセット

$\sigma_0 = \sqrt{T/(2\alpha\rho)}$ ▷ ガウシアンメカニズムの予算
$\epsilon_0 \leftarrow \sqrt{8(1-\alpha)\rho/T}$ ▷ 指数メカニズムの予算
$\rho_{\text{used}} \leftarrow 0$
$t \leftarrow 0$
Algorithm 3 を用いて θ_t を初期化.
for $r \in \mathcal{W}$ **do**
  $w_r = \sum_{s \in \mathcal{W}} c_s \mid r \cap s \mid$
while $\rho_{\text{used}} < \rho$ **do**
  $t \leftarrow t + 1$
  $\rho_{\text{used}} \leftarrow \rho_{\text{used}} + \frac{1}{8}\epsilon_t^2 + \frac{1}{2\sigma_t^2}$
  select 指数メカニズムを用いて $r_t \in \mathcal{W}_+$ を選択する：

$$q_r(D) = w_r \left(\|p_r(D) - p_r(\theta_{t-1})\|_1 - \sqrt{2/\pi} \cdot \sigma_t \cdot d_r \right)$$

  measure r_t に関する周辺分布にノイズを足し測定する：

$$\tilde{y}_t = p_{r_t}(D) + \mathcal{N}(0, \sigma_t^2 I)$$

  estimate Private-PGM を用いて分布を推定する：

$$\theta_t = \arg\min_{\theta \in S} \sum_{i=1}^{t} \frac{1}{\sigma_i} \|p_{r_i}(\theta) - \tilde{y}_i\|_2^2$$

  ϵ_{t+1} と σ_{t+1} を Algorithm 4 を用いて更新する
generate Private-PGM を用いて θ_t から合成データ $\hat{D}$ を生成.
return $\hat{D}$

Algorithm 3 初期化アルゴリズム

for $r \in \{r \in \mathcal{W}_+ \mid |r| = 1\}$ **do**
  $t \leftarrow t+1, \quad \sigma_t \leftarrow \sigma_0, \quad r_t \leftarrow r$ ▷ 1 次元周辺分布のみ学習
  $\tilde{y}_t = p_r(D) + \mathcal{N}(0, \sigma_t^2 I)$
  $\rho_{\text{used}} \leftarrow \rho_{\text{used}} + \frac{1}{2\sigma_t^2}$
$\theta_t = \arg\min_{\theta \in S} \sum_{i=1}^{t} \frac{1}{\sigma_i} \|p_{r_i}(\theta) - \tilde{y}_i\|_2^2$ ▷ Private-PGM

Algorithm 4 予算アニーリング

if $\|p_{r_t}(\theta_t) - p_{r_t}(\theta_{t-1})\|_1 \leq \sqrt{2/\pi} \cdot \sigma_t \cdot d_{r_t}$ **then** ▷ 更新の差分が小さい場合は予算増加
  $\epsilon_{t+1} \leftarrow 2 \cdot \epsilon_t$, $\sigma_{t+1} \leftarrow \sigma_t / 2$
else ▷ 更新の差分が小さくない場合は予算継続
  $\epsilon_{t+1} \leftarrow \epsilon_t$, $\sigma_{t+1} \leftarrow \sigma_t$
if $(\rho - \rho_{\text{used}}) \leq 2\left(\frac{1}{2\sigma_{t+1}^2} + \frac{1}{8}\epsilon_{t+1}^2\right)$ **then** ▷ 最後の使い切り
  $\epsilon_{t+1} = \sqrt{8 \cdot (1-\alpha) \cdot (\rho - \rho_{\text{used}})}$
  $\sigma_{t+1} = \sqrt{1/(2 \cdot \alpha \cdot (\rho - \rho_{\text{used}}))}$

が大きいときは予算をキープ．(6) 予算が尽きるまで (2) から (5) を繰り返す．

また，予算はすべて zCDP で用いる ρ に変換し，命題 2.7 の観点で ρ を単純に足していくことで管理している．本稿では，(1) の初期化と (3) のガウシアンメカニズムについてその有効性を検証する．

2.5 PrivBayes

PrivBayes は Zhang らが提案したベイジアンネットワークを用いた DP 合成手法である [17]．PrivBayes の学習は構造学習とパラメータ学習の大きく 2 段階に分けられる．特に，構造学習の際は，各属性をノードとみて関係性の強いノード間をエッジで結びグラフを作成する．関係性の強さは相互情報量で表現し，その大きさをスコアとした指数メカニズムを用いて結ぶエッジを選択する．

表 1 用いたデータセットのレコード数と属性の数と各属性の候補数.

Dataset	レコード数	属性数と各属性の候補数
Adult	30162	9 属性（8,16,7,14,6,5,2,41,2）
Bank	30488	11 属性（11,3,7,2,2,2,2,12,5,3,2）

3. 評価項目

AIM は提案論文にて著者らが各工夫の有効性を検証しているが，具体的な改善度合いを数値で評価しているわけではない [9]．本稿では，具体的な数値とともに評価する．実験で用いたコードは McKenna らによる実装である[*2][*3]．

3.1 評価目的

本稿の目的は，AIM の特徴である「初期化」「ガウシアンメカニズム」「zCDP による予算管理」の有効性を調査することである．また，生成された合成データが元データの属性間の相互情報量をどれくらい保持するかも検証する．本稿で取り組むの下記 3 点である．

- 生成された合成データの有用性評価により「初期化」「ガウシアンメカニズム」の有効性を確かめる
- zCDP による予算管理の有効性を確かめる
- 合成データが相互情報量を保持するかを確かめる

3.2 合成データ評価方法

評価実験には 2 つのデータセットを用いた．Adult [1] と Bank [14] である．それぞれカテゴリカルな属性のみを使い，欠損値があるものは削除した．扱ったデータセットのレコード数，属性数は表 1 の通りである．

合成データの有用性評価は，α-way 誤差（$\alpha = 1, 2, 3$）といった統計的類似性と，機械学習モデルの精度である．α-way 誤差はワークロード誤差の特殊形として定義できる．ワークロード誤差は下記の通りである．

定義 3.1 (ワークロード誤差, α-way 誤差). $\mathcal{W} := \{r_1, \ldots, r_k\}$ を測定対象，$c_1, \ldots, c_k \geq 0$ を重みとする．元データ D と合成データ $\hat{D}$ の**ワークロード誤差**は次の通りである

$$L(D, \hat{D}) := \frac{1}{k \cdot |D|} \sum_{i=1}^{k} c_i \|p_{r_i}(D) - p_{r_i}(\hat{D})\|_1.$$

ここで，$W(n) := \{r \in 2^{[d]} \mid |r| = n\}$ と定める．特に，す

[*2] https://github.com/ryan112358/mbi
[*3] 投稿直前の時期，2025 年 8 月 10 日ごろ，リポジトリ名やいくつかのファイルに変更や更新があった．本稿で用いたコミット ID は「1b0d4363e1c4de04de2a4ef5cfa441c8a82910ee」である．

べての $c_i = 1$ としたとき，$W(\alpha)$ に対するワークロード誤差を **α-way 誤差**と呼ぶこととする．

機械学習モデルの精度については合成データで学習したものの評価をした．決定木，サポートベクトルマシン，XGBoost の 3 つモデルを用いて，Adult は income，Bank は y という変数の 2 値分類を行った．

最後に元データ D の属性 A_i, A_j の間の相互情報量と合成データ $\hat{D}$ の属性 A_i, A_j の間の相互情報量を比較する．2 つの確率変数 X, Y の相互情報量は次のように定義される．

$$I(X,Y) := \sum_{\substack{x \in \mathrm{dom}(X),\\ y \in \mathrm{dom}(Y)}} \Pr[(X,Y)=(x,y)] \log \frac{\Pr[(X,Y)=(x,y)]}{\Pr[X=x]\Pr[Y=y]}.$$

X, Y が独立のときは $I(X,Y) = 0$ となり，関係性が強いときに大きい値になる．

3.3 評価対象の手法

下記 5 種類の手法を比較する．既存実装のデフォルト設定に従い，$\varepsilon = 1$，$\delta = 10^{-9}$ とした．

- `aim-k2`：測定対象を $W(2)$ としたときの AIM．デフォルト設定だとこれになっている．
- `aim-k3`：測定対象を $W(3)$ としたときの AIM．
- `aim-noinit`：初期化アルゴリズムを適用しなかった AIM．測定対象は $W(2)$．
- `aim-lap`：ガウシアンメカニズムをラプラスメカニズムに変えたもの．測定対象は $W(2)$．
- `bayes`：PrivBayes [17]（親ノード候補数は 2）．実装は DataSynthesizer [15] を用いた．

4. 評価実験結果

3.1 項で列挙した点を 1 点ずつ調査していく．

4.1 各結果の有用性について

まず，各手法についての α-way 誤差の結果は図 1, 2 の通りになった．それぞれ 3 回合成を行い，ワークロード誤差の平均値を棒グラフで表現している．標準偏差をエラーバーとして記載している．また，ML 有用性についても表 2, 3 の通りになった．

どちらの結果も共通して，AIM が PrivBayes より優れた結果となった．また，AIM の手法内では `aim-2k` が優れた結果になり，ガウシアンメカニズムをラプラスメカニズムを取り換えた `aim-lap` が最も悪い結果になった．初期化を行わない `aim-noinit` は平均値という観点では Adult ではわずかに `aim-2k` より悪い結果になったが，エラーバーの大きさも加味すると差があるとは言えない結果になった．また，Bank の結果はエラーバーだけでなく平均値も含めてあまり違いのない結果となった．

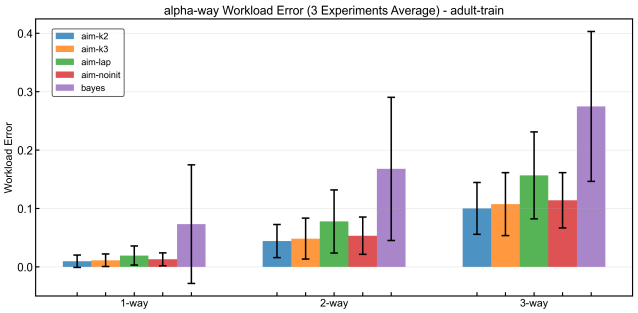

図 1 Adult に対する各手法による合成データの α-way 誤差．

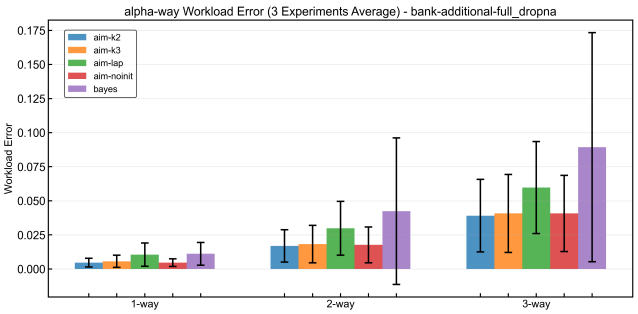

図 2 Bank に対する各手法による合成データの α-way 誤差．

4.2 ステップごとの結果

AIM は適応的に予算の消費ペースを変えていく．各ステップでのプライバシー損失予算やデータの有用性，アルゴリズムの挙動などを可視化した．全ての可視化はデータセットが Adult の場合のみを対象としている．

4.2.1 各ステップでのプライバシー損失予算の推移

各ステップで消費されているプライバシー損失予算の推移を図 3 にまとめた．zCDP が AIM らで用いられた zCDP による予算管理で，prvacc は数値的合成則による予算管理である．$\varepsilon = 1, 5, \delta = 10^{-5}$ の場合について検証した．数値的合成則は ε を近似計算するがその上限と下限を水色の領域で表している．

全ての手法について，2 回ほど上昇のペースが変わっている場所がある．そこで Algorithm 4 での予算増加が行われている（1 行目の If 文に入っている）ことがわかる．

結果から zCDP による予算管理は数値的合成則の結果とほとんど変わらない結果となった．これは，AIM における zCDP による予算管理は，数値的合成則の観点でもほとんど最適であることを表している．

4.2.2 各ステップでの α-way 誤差の大きさ

各ステップでの `aim-2k` の α-way 誤差を図 4 に記した．最初の 9 ステップは初期化部分（Algorithm 3）であるので，9 ステップと 10 ステップの間に点線を引いた．

特に，1-way 誤差については初期化アルゴリズムが終わった段階で大きく誤差が減っていることが見られる．しかしながら，図 1 において，`aim-noinit` が `aim-2k` とほとんど変わらなかったことより，初期化をしない場合でも

表 2　Adult に対する各モデルの Classification Accuracy

Algorithm	aim-k2	aim-k3	aim-lap	aim-noinit	bayes
決定木	**0.8190** ± 0.0005	0.8188 ± 0.0013	0.8048 ± 0.0224	0.8157 ± 0.0026	0.8021 ± 0.0224
Support vector machine	**0.8183** ± 0.0025	0.8166 ± 0.0010	0.8040 ± 0.0251	0.8160 ± 0.0004	0.8034 ± 0.0233
XGBoost	**0.8208** ± 0.0021	0.8173 ± 0.0025	0.8036 ± 0.0231	0.8176 ± 0.0017	0.8029 ± 0.0214

表 3　Bank に対する各モデルの Classification Accuracy

Algorithm	aim-k2	aim-k3	aim-lap	aim-noinit	bayes
決定木	**0.8841** ± 0.0008	0.8821 ± 0.0008	0.8787 ± 0.0062	0.8833 ± 0.0016	0.8811 ± 0.0079
Support vector machine	**0.8847** ± 0.0002	0.8821 ± 0.0009	0.8784 ± 0.0064	0.8846 ± 0.0011	0.8811 ± 0.0077
XGBoost	**0.8841** ± 0.0004	0.8819 ± 0.0006	0.8782 ± 0.0058	**0.8841** ± 0.0011	0.8808 ± 0.0072

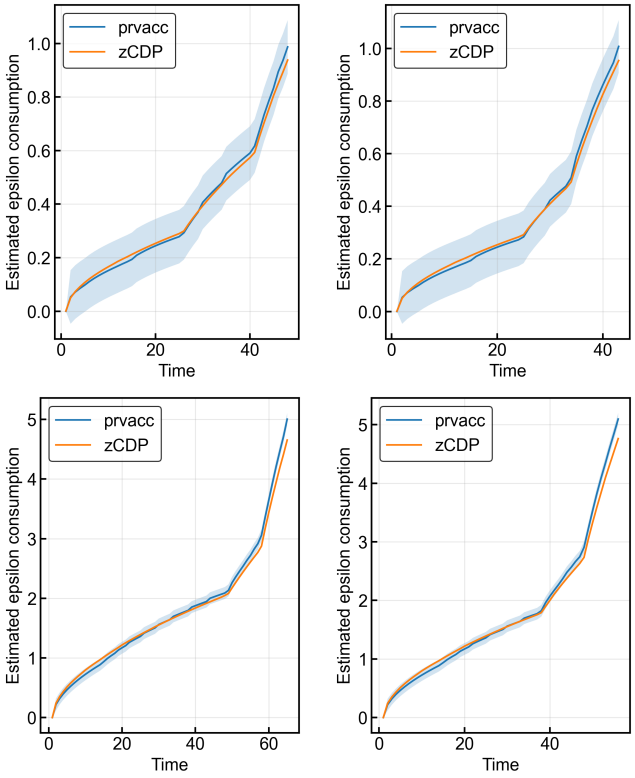

図 3　AIM の各ステップにおける予算消費の様子．上段が $\varepsilon = 1$ の場合，下段が $\varepsilon = 5$ の場合．それぞれ $\delta = 10^{-5}$．左から $W(2), W(3)$ の場面．zCDP による予算管理と数値的合成則（prvacc）による予算管理がほとんど変わらないことがわかる．また，全体として，2 回ほど予算消費ペースが変わった箇所があることが確かめられた．

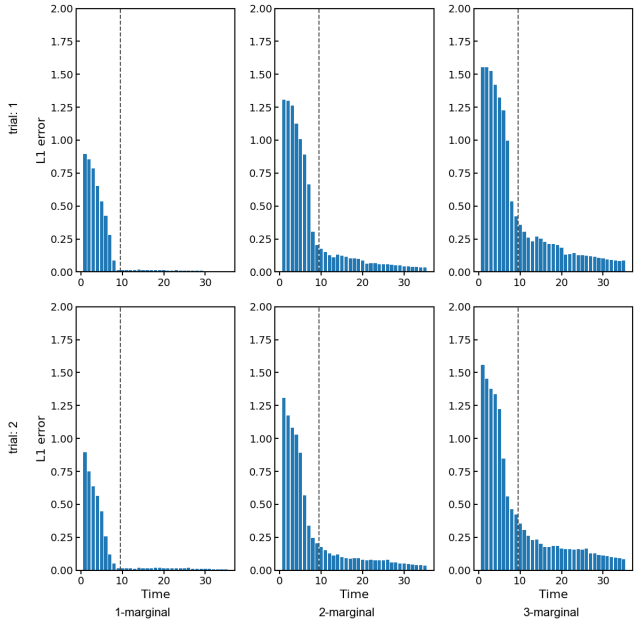

図 4　Adult の aim-2k における各ステップでの α-way 誤差の大きさ．試行は 2 回行い，上段が 1 回目，下段が 2 回目である．

のちの最適化で遅れが取り返せるということもわかる．

4.2.3　各ステップで選ばれた属性の組

3 回の試行について aim-2k において，各ステップで選ばれた属性の組を図 5 に記した．また，その際に選ばれた属性の組のみのワークロード誤差と選ばれなかった属性の組のみのワークロード誤差も同様にグラフにした．

3 回とも異なる結果ではあるが，選ばれやすい属性と選ばれにくい属性が見られた．また，選ばれた属性の L1 誤差の平均と選ばれなかった属性の L1 誤差の平均を比較すると，やはり選ばれたものの方が小さくなっていることが確認できた．

4.3　相互情報量の保存について

各属性の組について，元データでの相互情報量と合成データでの相互情報量を比較する．平均的な相互情報量のずれの大きさが図 8 である．AIM が PrivBayes より大幅に平均的な相互情報量も保存していることが確かめられた．aim-2k と aim-noinit は同じくらいの結果であった．

相互情報量の大きさをヒートマップで表したのが，図 6 と図 7 である．全ての手法で特に相互情報量の大きい属性の組を保存していることがわかる．

5.　制限と今後の課題

本稿の報告は $\varepsilon = 1$ の結果を中心に行った．ε を大きい値にすると，鏡像降下法の学習が終わらなくなってしまう問題が確認できたからである．具体的には通常の学習・データ合成が 1 時間程度で終わるのに対して，2 日以上まっても処理が終わらないなどの現象が確認できた．これらは計算の高速化のためとみられる追加実装に関係する箇所で発生しており，バージョンを下げたりすることで回避するこ

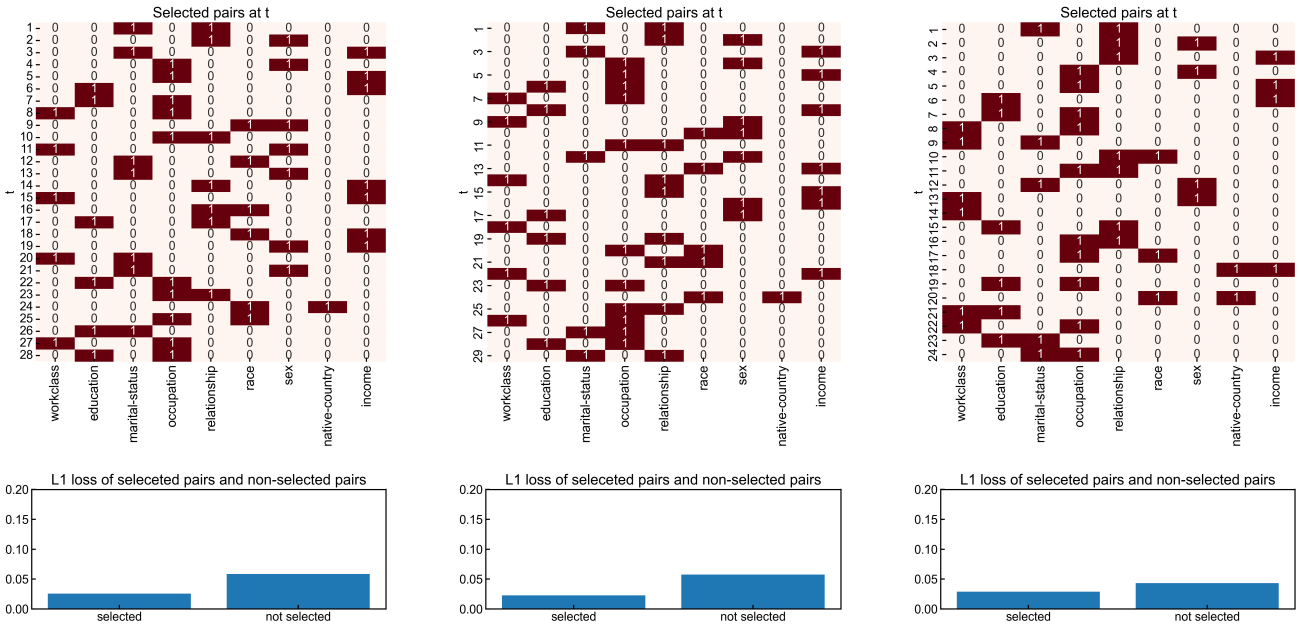

図 5　$\varepsilon = 1, \delta = 10^{-5}$ のときの aim-2k において各ステップで指数メカニズムに選択された属性の組．色が濃い場所が選ばれた属性．選択された属性組と選択されなかった属性組の L1 誤差平均．選ばれた属性の平均の方が小さいことがわかる．

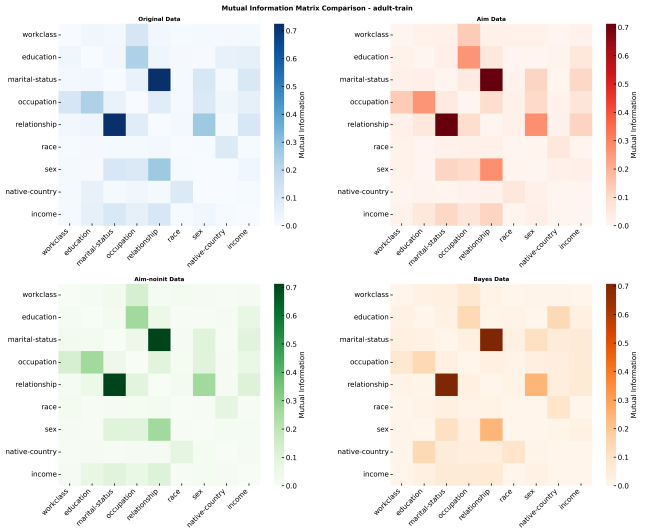

図 6　Adult の各属性の組の相互情報量の大きさ．左上が元データ．右上が aim-2k．左下が aim-noinit．右下が PrivBayes.

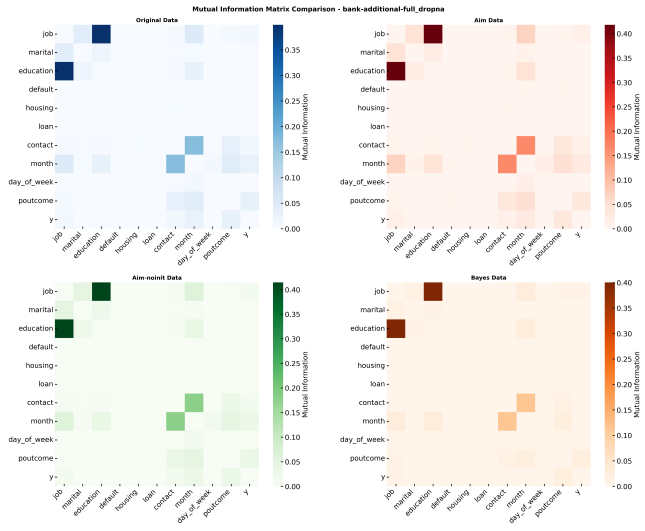

図 7　Bank の各属性の組の相互情報量の大きさ．左上が元データ．右上が aim-2k．左下が aim-noinit．右下が PrivBayes.

ともできるが，より正確な解決・回避は今後の課題とする．

6.　まとめ

本稿では，AIM が優れた品質のデータを生成する要因分析を行った．グラフィカルモデルの定式化や内部関数の Private-PGM，AIM のアルゴリズムをそれぞれ解説しつつ，初期化の工夫やノイズ付加アルゴリズムの有効性を調査した．実験の結果，陽に属性間の相互情報量を学習する PrivBayes よりも，AIM の方が元データの属性間の相互情報量を保存していることを確かめた．また，AIM は zCDP を用いて予算管理を行っていたが，それを数値的合成則の

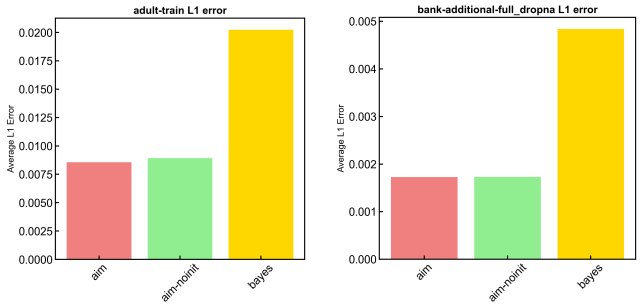

図 8　元データと合成データのそれぞれの各属性間の相互情報量の L1 誤差．左が Adult，右が Bank の結果

観点に焼き直してもタイトな予算管理を行えていることを明らかにした．

参考文献

[1] Becker, B. and Kohavi, R.: Adult, UCI Machine Learning Repository (1996). DOI: https://doi.org/10.24432/C5XW20.

[2] Bun, M. and Steinke, T.: Concentrated differential privacy: Simplifications, extensions, and lower bounds, *Theory of Cryptography Conference*, Springer, pp. 635–658 (2016).

[3] Chen, K., Li, X., Gong, C., McKenna, R. and Wang, T.: Benchmarking differentially private tabular data synthesis, *arXiv preprint arXiv:2504.14061* (2025).

[4] Dwork, C.: Differential privacy, *International Colloquium on Automata, Languages, and Programming*, Springer, pp. 1–12 (2006).

[5] Gopi, S., Lee, Y. T. and Wutschitz, L.: Numerical composition of differential privacy, *Advances in Neural Information Processing Systems*, Vol. 34, pp. 11631–11642 (2021).

[6] Kairouz, P., Oh, S. and Viswanath, P.: The composition theorem for differential privacy, *International conference on machine learning*, PMLR, pp. 1376–1385 (2015).

[7] Li, H., Xiong, L., Zhang, L. and Jiang, X.: DPSynthesizer: Differentially private data synthesizer for privacy preserving data sharing, *Proceedings of the VLDB Endowment International Conference on Very Large Data Bases*, Vol. 7, No. 13, NIH Public Access, p. 1677 (2014).

[8] Liew, S. P., Takahashi, T. and Ueno, M.: PEARL: Data Synthesis via Private Embeddings and Adversarial Reconstruction Learning, *International Conference on Learning Representations* (2022).

[9] McKenna, R., Mullins, B., Sheldon, D. and Miklau, G.: Aim: An adaptive and iterative mechanism for differentially private synthetic data, *arXiv preprint arXiv:2201.12677* (2022).

[10] McKenna, R., Sheldon, D. and Miklau, G.: Graphical-model based estimation and inference for differential privacy, *International Conference on Machine Learning*, PMLR, pp. 4435–4444 (2019).

[11] McSherry, F. and Talwar, K.: Mechanism design via differential privacy, *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, IEEE, pp. 94–103 (2007).

[12] Mironov, I.: Rényi differential privacy, *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, IEEE, pp. 263–275 (2017).

[13] Miura, T., Kimura, E., Ichikawa, A., Kii, M. and Yamamoto, J.: Evaluating Synthetic Data Generation Techniques for Medical Dataset, *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 2: HEALTHINF*, INSTICC, SciTePress, pp. 315–322 (online), DOI: 10.5220/0012314500003657 (2024).

[14] Moro, S., R. P. and Cortez, P.: Bank Marketing, UCI Machine Learning Repository (2014). DOI: https://doi.org/10.24432/C5K306.

[15] Ping, H., Stoyanovich, J. and Howe, B.: Datasynthesizer: Privacy-preserving synthetic datasets, *Proceedings of the 29th International Conference on Scientific and Statistical Database Management*, pp. 1–5 (2017).

[16] Wainwright, M. J., Jordan, M. I. et al.: Graphical models, exponential families, and variational inference, *Foundations and Trends® in Machine Learning*, Vol. 1, No. 1–2, pp. 1–305 (2008).

[17] Zhang, J., Cormode, G., Procopiuc, C. M., Srivastava, D. and Xiao, X.: PrivBayes: Private Data Release via Bayesian Networks, *ACM Trans. Database Syst.*, Vol. 42, No. 4 (online), DOI: 10.1145/3134428 (2017).

[18] Zhong, Y., Ge, Y., Qin, J., Zheng, S., Tang, B., Qiu, Y.-X., Mao, R., Yuan, Y., Onizuka, M. and Xiao, C.: Privacy-Enhanced Database Synthesis for Benchmark Publishing, *arXiv preprint arXiv:2405.01312* (2024).

付　　録

A.1 鏡像降下法について

グラフィカルモデル $\theta \in \mathbb{R}^n$ を鏡像降下法で学習する方法について説明する．全ての処理は並列で行えるので，1つの測定対象について行われていると考える．周辺分布は

$$\mu = \frac{1}{Z}\exp\theta \in \mathbb{R}^n$$

である．いま，計算したいのは

$$\operatorname*{argmin}_{\mu} L(\mu)$$

である．鏡像降下法とは，

$$\mu^{t+1} = \operatorname*{argmin}_{\mu\in\mathcal{M}} \left(\mu^\top \nabla L(\mu^t) + \frac{1}{\eta_t} D_B(\mu, \mu^t)\right)$$

という更新方法で, 反復的に目的関数を最小化する μ^t を計算する学習方法である．ここで，D_B は Bregman divergence であり，狭義凸関数 $\psi : \mathbb{R}^d \to \mathbb{R}$ を用いて

$$D_B(\mu_1, \mu_2) := \psi(\mu_1) - \psi(\mu_2) - (\mu_1 - \mu_2)^\top \nabla_\mu \psi(\mu_2)$$

と定義される関数である．また，本稿では凸関数としてエントロピー

$$\psi(\mu) := -H(\mu) = \sum_i \mu_i \log \mu_i$$

を用いることとする．ここで，

$$(\nabla_\mu H(\mu))_i = -\log\mu_i - 1 = -\theta_i - 1 + \log Z$$

である．次の式変形から，グラフィカルモデルについては，μ ではなく θ を更新すればいいことがわかる．

$$\begin{aligned}
\mu^{t+1} &= \operatorname*{argmin}_{\mu\in\mathcal{M}} \mu^T \nabla_\mu L(\mu^t) + \frac{1}{\eta_t} D_B(\mu, \mu^t) \\
&= \operatorname*{argmin}_{\mu\in\mathcal{M}} \mu^T \nabla_\mu L(\mu^t) + \frac{1}{\eta_t}\left(-H(\mu) + \mu^T \nabla_\mu H(\mu^t)\right) \\
&= \operatorname*{argmin}_{\mu\in\mathcal{M}} \mu^T \left(\nabla_\mu L(\mu^t) + \frac{1}{\eta_t}\nabla_\mu H(\mu^t)\right) - \frac{1}{\eta_t} H(\mu) \\
&= \operatorname*{argmin}_{\mu\in\mathcal{M}} \mu^T \left(\eta_t \nabla_\mu L(\mu^t) + \nabla_\mu H(\mu^t)\right) - H(\mu) \\
&= \operatorname*{argmin}_{\mu\in\mathcal{M}} \mu^T \left(\eta_t \nabla_\mu L(\mu^t) - \theta^t\right) - H(\mu) \\
&= \mu\left(\theta^t - \eta_t \nabla_\mu L(\mu^t)\right)
\end{aligned}$$

最後の式変形は文献 [16] に従う．