多値時系列データの遷移情報を用いた局所差分プライバシー手法 Transition RRの提案

山本 悠太[1,a)]　牛田 芽生恵[1,2]　菊池 浩明[3]

概要：継続的に収集される時系列データから利用状況の分布を推定する需要は高い一方，そのデータには個人のセンシティブな情報が含まれうる．局所差分プライバシー (LDP) は各ユーザが手元でデータを摂動して送信する枠組みであり，信頼できる収集者を前提としない点で有力である．しかし継続的な収集では逐次合成により全時刻へ予算を分割する必要が生じ，素朴な手法では時間長の増加とともに精度が著しく劣化する．本稿では，視聴率推定を想定した多値カテゴリの時系列頻度推定を対象とし，時刻間の状態遷移そのものをランダム化応答の入力とすることで，変化のない時刻には予算を消費しない Transition RR を提案する．これを，時間分割型の一般化ランダム化応答 (GRR)，時間・チャンネル分割型のランダム化応答，無変化時に予算を消費しない差分報告機構 DDRM の多値拡張と同一の枠組みで比較し，プライバシー予算・チャンネル数・時間長に対する推定誤差を報告する．実験により，Transition RR はチャンネル数が小さく時間長が中程度以下で予算が中程度の領域において最も高精度である一方，チャンネル数が大きい領域や時間長が長い領域では DDRM が優位となることを示す．

キーワード：局所差分プライバシー，時系列データ，頻度推定，ランダム化応答，視聴率

Proposal a Local Differential Privacy Method for Multivalued Time Series Data *Transition RR*

Yuta Yamamoto[1,a)]　Mebae Ushida[1,2]　Hiroaki Kikuchi[3]

Abstract: Estimating the distribution of user activity from continuously collected time-series data is widely demanded, yet such data may contain sensitive personal information. Local differential privacy (LDP) is a promising framework in which each user perturbs data locally before submission, requiring no trusted collector. In continual collection, however, sequential composition forces the privacy budget to be divided across all timestamps, and naive approaches degrade sharply as the time horizon grows. This paper targets multi-valued time-series frequency estimation for audience-rating estimation, and proposes Transition RR, which feeds the state transition between consecutive timestamps into randomized response and consumes no budget when the state is unchanged. We compare it with time-split generalized randomized response (GRR), time-and-channel-split randomized response, and the multi-valued extension of the Dynamic Difference Report Mechanism (DDRM) in a unified framework, reporting estimation errors with respect to the privacy budget, the number of channels, and the time horizon. Experiments show that Transition RR achieves the lowest error when the number of channels is small, the time horizon is moderate, and the budget is moderate, whereas DDRM is superior when the number of channels is large or the time horizon is long.

Keywords: local differential privacy, time-series data, frequency estimation, randomized response, audience rating

[1] 明治大学大学院 Meiji University graduate school
[2] 富士通 Fujitsu Ltd.
[3] 明治大学 Meiji University
[a)] cs252046@meiji.ac.jp

1. 背景

デジタル化に伴い個人データの収集・分析が広がる一方，

© 2026 Information Processing Society of Japan

This paper is work in progress and not peer-reviewed.

プライバシー侵害のリスクも増大している．その対策として，データの有用性を保ちつつ各ユーザのプライバシーを保護して統計量を求める LDP (Local Differential Privacy，局所差分プライバシー) [1], [2] が注目されている．LDP では各ユーザが手元でデータを摂動してから送信するため，信頼できる収集者を前提としない．

本研究が想定する具体的用途は，テレビや動画配信における視聴率の継続的推定である．各ユーザは「チャンネルごとに視聴中か否か」を表す二値をチャンネル数ぶん保持し，一定時刻ごとに送信するため，収集対象は「時系列かつ多次元（チャンネル数次元)」のデータとなる．このデータには，(1) ある瞬間に視聴中のチャンネルは高々一つであること，(2) 視聴チャンネルは頻繁には変化しないこと，という二つの重要な性質がある．

既存の LDP 研究の多くは一時点の収集を前提としており，継続収集にそのまま適用すると逐次合成定理により予算が時刻数で分割され，精度が時間とともに劣化する．継続収集向けの LDP 手法として，memoization に基づく RAPPOR [3] や dBitFlipPM [4]，Xue らによる時系列の変化点に着目する差分報告機構 DDRM [5] が提案されている．中でも DDRM は「値が変化しない時刻には予算を消費しない」性質により継続収集に適するが，**二値時系列を対象としており多値データを直接には扱えない**という欠点がある．その多値拡張 [5] は，one-hot 符号化したうえでユーザを成分ごとのグループに分割し各自が 1 ビットのみを報告する，すなわち多値を二値へ帰着させる方式であり，各成分の推定に使えるユーザ数が $1/C$ に減るうえ，「同時に視聴できるチャンネルは高々一つ」という排他性を摂動の設計に活かせない．

そこで本研究では，多値データを二値へ帰着させるのではなく，**時刻間の状態遷移そのものをランダム化応答の入力とする**方式 Transition RR を提案する．遷移を報告単位とすることで，変化のない時刻には入力に依存しない一様な出力を返せるため予算を消費せず，ユーザ分割が不要なため全 n 人を全チャンネルの推定に使え，さらに排他性を遷移集合の定義域として自然に取り込める点が新規性である．

本稿の貢献は，(1) 多値時系列を二値へ帰着させずに扱う Transition RR を提案し ε-LDP と不偏性を示すこと，(2) Transition RR の ℓ_2 誤差上界に加え，DDRM 多値拡張の誤差上界を導出し，各方式が優位となる条件を T と C の関数として明らかにすること，(3) 4 手法を同一の枠組みで実装した合成データ実験により，その条件を裏づけること，の三点である．

2. 問題

2.1 データモデル

n 人のユーザ $u_1, \ldots, u_n$，C 個のチャンネル，T 個の時刻を考える．ユーザ u_i の時刻 t における視聴状況を二値ベクトル $\boldsymbol{b}_i^t = (b_{i,1}^t, \ldots, b_{i,C}^t) \in \{0,1\}^C$ で表し，$b_{i,c}^t = 1$ はチャンネル c を視聴中であることを表す．すなわち収集対象は $n \times C \times T$ の二値テンソルである．

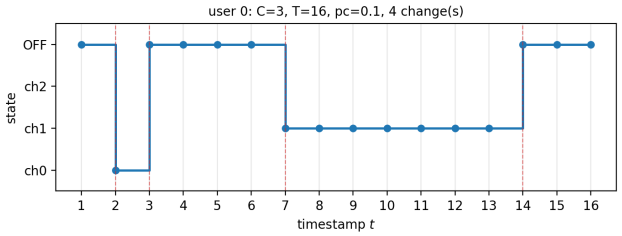

図 1 1 ユーザの視聴状況の遷移例（$C = 3$，$T = 16$，$p_c = 0.1$）．縦軸は状態，破線は変化点を表す．

視聴データには次の性質を仮定する．

相互排他性 各時刻に視聴中のチャンネルは高々一つ，すなわち $\sum_c b_{i,c}^t \le 1$ である．非視聴状態，すなわち全成分が 0 の場合を $C+1$ 番目のチャンネルとみなす．この性質により，$\boldsymbol{b}_i^t$ は $M = C+1$ 値の状態 $x_i^t \in \mathcal{X} = \{1, \ldots, C, \mathrm{OFF}\}$ と一対一に対応し，各時刻の状態は $\mathcal{X}$ 上の 1 点として一意に定まる．

低い変化率 連続する時刻で状態が変化する確率 $p_c = \Pr[x_i^t \neq x_i^{t-1}]$ は小さい．視聴者は同一チャンネルを一定時間視聴し続けるためである．

図 1 に 1 ユーザの視聴状況の遷移例を示す．16 時刻のうち状態が変化するのは 4 時刻のみであり，大半の時刻では前時刻と同じ状態が続いている．Transition RR および DDRM は，この「変化のない時刻」に予算を消費しないことで継続収集に対応する．

推定対象とプライバシー要件について，収集者は各時刻 t において，チャンネルごとの視聴率

$$f_c^t = \frac{1}{n} \sum_{i=1}^{n} b_{i,c}^t \qquad (c = 1, \ldots, C) \qquad (1)$$

を推定したい．すなわち推定対象は $C \times T$ の頻度行列 $\boldsymbol{f}^t = (f_1^t, \ldots, f_C^t)$ である．

収集者は信頼できないものとし，各ユーザは自身のデータを手元で摂動してから送信する．プライバシー保証は各ユーザの時系列全体に対する ε-LDP（user-level）とする．

定義 1 (ε-LDP [1])．入力の定義域を $\mathcal{D}$，出力の定義域を $\mathcal{V}$ とする．ランダム化アルゴリズム $\mathcal{A} : \mathcal{D} \to \mathcal{V}$ が *ε-LDP* を満たすとは，任意の入力 $d, d' \in \mathcal{D}$ と任意の出力 $v \in \mathcal{V}$ に対して $\Pr[\mathcal{A}(d) = v] \le e^{\varepsilon} \Pr[\mathcal{A}(d') = v]$ が成り立つことをいう．

本稿では入力の定義域は $\mathcal{D} = \mathcal{X}^T$ で全方式共通であり，出力の定義域 $\mathcal{V}$ のみが方式ごとに異なる（RR は $(\{0,1\}^C)^T$，GRR は $\mathcal{X}^T$，Transition RR は遷移集合 $\mathcal{T}$ を用いて $\mathcal{T}^T$）．

© 2026 Information Processing Society of Japan

表 1　トイデータ

ユーザ	$t=1$	$t=2$	$t=3$
u_1	A	A	B
u_2	OFF	A	A
u_3	C	C	C

LDP は逐次合成性をもち，$\varepsilon_1,\ldots,\varepsilon_S$-LDP を満たす S 個のアルゴリズムを同一データに順に適用した出力全体は $\sum_s \varepsilon_s$-LDP を満たす．したがって T 時刻すべてで報告を行う素朴な方式では，各時刻に ε/T しか割り当てられない．

推定精度は，全時刻・全チャンネルにわたる ℓ_2 誤差

$$\ell_2 = \sqrt{\sum_{t=1}^{T}\sum_{c=1}^{C}\left(\hat{f}_c^t - f_c^t\right)^2} \tag{2}$$

で評価する．

2.2　トイデータ

以降の説明では，**表 1** に示す 3 ユーザ・3 チャンネル $(\mathrm{A},\mathrm{B},\mathrm{C})$・3 時刻のトイデータを共通に用いる．状態数は $M=C+1=4$ である．真の視聴率は $\boldsymbol{f}^1=(\frac{1}{3},0,\frac{1}{3})$，$\boldsymbol{f}^2=(\frac{2}{3},0,\frac{1}{3})$，$\boldsymbol{f}^3=(\frac{1}{3},\frac{1}{3},\frac{1}{3})$ となる．

3.　先行研究

3.1　RR と GRR

ランダム化応答 (RR) [6] は LDP の最も基本的な機構である．二値 $x\in\{0,1\}$ に対し，真の値を確率 $p=e^\varepsilon/(e^\varepsilon+1)$ で，反転した値を確率 $1-p$ で出力すれば ε-LDP を満たす．出力の平均 $\bar{y}$ から $\hat{f}=(\bar{y}-(1-p))/(2p-1)$ により 1 の頻度を不偏に推定できる．

一般化ランダム化応答 (GRR) [7], [8] は RR を M 値のカテゴリ $x\in\{1,\ldots,M\}$ に拡張したもので，

$$\Pr[Y=y\mid x]=\begin{cases} p=\dfrac{e^\varepsilon}{e^\varepsilon+M-1}, & y=x,\\[2mm] q=\dfrac{1}{e^\varepsilon+M-1}, & y\neq x,\end{cases} \tag{3}$$

とする．値 j を報告したユーザ数を n_j とすると，$\hat{f}_j=(n_j/n-q)/(p-q)$ は不偏推定量であり，その分散は M に比例して増加する．

3.1.1　本問題への適用

一時点向けの RR/GRR を継続収集に適用する最も素朴な方法は，逐次合成により予算を分割することである．本稿では次の 2 通りをベースラインとする．

3.1.2　GRR（時間分割）

各時刻に状態 $x_i^t\in\{\mathrm{A},\mathrm{B},\mathrm{C},\mathrm{OFF}\}$ を $M=C+1$ 値の GRR で報告する．各報告の予算は ε/T である．トイデータでは $M=4$，$T=3$ であり，各報告に $\varepsilon/3$ が割り当てられる．たとえば u_1 は $t=1,2,3$ にそれぞれ真値 A，A，B を入力とし，確率 $p=e^{\varepsilon/3}/(e^{\varepsilon/3}+3)$ で真値を，そうでなければ他の 3 状態から一様に選んだ値を報告する．u_1 の報告が (A, C, B) となった場合，$t=2$ の C は摂動により生じた偽の報告である．収集者は各時刻の報告数から $\hat{f}_j^t=(n_j^t/n-q)/(p-q)$ を計算する．この方式は相互排他性を GRR の定義域として自然に取り込めるが，予算が T 等分されるため長期の収集で精度が劣化する．

3.1.3　RR（時間・チャンネル分割）

各時刻に各チャンネルの視聴状況をそれぞれ報告する．1 ユーザが各時刻 2 値 RR を C 回行い報告するため，各報告の予算は $\varepsilon/(TC)$ である．トイデータでは 1 ユーザが計 9 回報告し，各報告に $\varepsilon/9$ が割り当てられる．この方式は相互排他性を利用しないため，「A と C を同時に視聴中」のような真には起こり得ない報告が生じ，予算の分割数も C 倍になる．

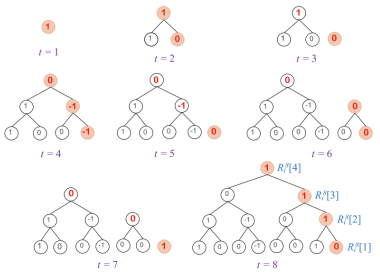

図 2　差分木の更新手順（$d_i=[1,1,1,0,0,0,1,1]$，文献 [5] をもとに作成）．橙色は現在値に関係するノードを表す．
Fig. 2　Update procedure of the difference trees.

3.2　DDRM

DDRM (Dynamic Difference Report Mechanism) [5] は二値時系列 $d_i^t\in\{0,1\}$ に対する継続頻度推定機構である．主な構成要素は次の 4 つである．

3.2.1　差分木

各ユーザは連続する時刻の差分 $c_i^t=d_i^t-d_i^{t-1}\in\{-1,0,+1\}$ を葉とし，親が子の和をもつ完全二分木（差分木）を保持する．時刻 t を $t=2^{a_1}+\cdots+2^{a_m}$ と 2 進分解すると m 本の木が構成され，最右の木の根は $d_i^t-d_i^{t-2^{a_m}}$ という「複数時刻にわたる変化」を表す．ユーザは各レベルの最右ノードのみを保持すればよく，記憶量は $O(\log t)$ である．

図 2 に $d_i=[1,1,1,0,0,0,1,1]$ の場合の更新手順を示す．各時刻で新たな葉（差分 c_i^t）が右端に追加され，木を完全にする内部ノードが同時に確定する．橙色が現在値に関係するノードで，$t=8$ では葉 $R_i^8[1]$ から根 $R_i^8[4]$ の 4 つが該当する．$t=6=2^2+2^1$ のように 2 の冪でない場合は木が 2 本に分かれ，最右の木の根は $d_i^6-d_i^4$ となる．

3.2.2　ノード選択

各時刻，ユーザは葉（1 時刻の差分）と現在値に対応する最上位ノード（根）のいずれかを時刻に応じて確率的に

© 2026 Information Processing Society of Japan

選択し報告する．葉のみを報告すると推定誤差が時間とともに累積するが，根は $\hat{\boldsymbol{f}}^{t-2^a m}$ からの差分を与えるため，両者を分散に応じて重み付き平均することで累積誤差が抑制される．

3.2.3 予算配分

各ユーザは非ゼロ値の報告を高々 k 回に制限し，各報告に $\alpha = \varepsilon/k$ を割り当てる．k が小さいと予算を使い切ったユーザの誤報告（manipulation error）が，大きいと 1 報告あたりのノイズが支配的になるため，両誤差の上界が交差する点として予想報告回数 k が導出される．

3.2.4 摂動

選択したノードの値 $v \in \{-1, 0, +1\}$ に対し，$v \neq 0$ なら $\tilde{v} = v$ を確率 $e^\alpha/(e^\alpha + 1)$ で，$-v$ を残りの確率で出力する．$v = 0$ なら ± 1 を等確率で出力し，予算を消費しない．収集者は $\hat{v} = \tilde{v}\,(e^\alpha + 1)/(e^\alpha - 1)$ により校正し，$\mathbb{E}[\hat{v}] = v$ を得る．

3.2.5 多値への拡張:1-hot DDRM

DDRM は二値時系列を対象とするため，多値データには 1-hot 符号化した各ビットに DDRM を適用し, ユーザ全体を C グループに分割し，第 c グループのユーザはチャンネル c のビット $b^t_{i,c}$ のみを DDRM で報告するという方針が示されている [5]．これにより通信量と予算消費が抑えられるが, 各チャンネルの推定に使えるユーザ数は n/C に減少し，グループ分割による標本誤差が生じる．この方針を本稿では 1-hot DDRM と呼び, 以後の実験で使用する．

定理 1 (1-hot DDRM の ℓ_2 誤差上界)**.** 予算枯渇が生じないとき，ユーザを C グループに分割する *1-hot DDRM* の ℓ_2 誤差は,

$$\sqrt{\sum_{c=1}^{C} \left(\hat{f}^t_c - f^t_c\right)^2} \le \frac{2G_\alpha C}{\sqrt{n}}, \qquad \ell_2 \le 2G_\alpha C\sqrt{\frac{T}{n}} \quad (4)$$

を満たす（$G_\alpha = (e^\alpha + 1)/(e^\alpha - 1)$，$\alpha = \varepsilon/k$）．

すなわち ℓ_2 誤差は T に対して $\sqrt{T}$ のオーダーに留まる一方，係数は C に比例して増加する．

3.2.6 本問題への適用

トイデータでは 3 ユーザを 3 グループに分け，u_1 がチャンネル A，u_2 が B，u_3 が C を担当するとする．**表 2** に各ユーザの担当ビットの系列と差分，および差分木のノード値を示す．$t = 2$ では $2 = 2^1$ より高さ 2 の木が 1 本でき，その根は $c^1 + c^2 = d^2 - d^0$ となる．$t = 3$ では $3 = 2^1 + 2^0$ より 2 本の木ができ，最右の木は葉のみである．

$t = 1$ では葉のみが報告対象であり，u_1, u_3 は $v = +1$ を予算 α で摂動し，u_2 は $v = 0$ を予算消費なしで ± 1 に置き換える．$t = 2$ では各ユーザが葉 ($v = 0$) と根 (u_1, u_3 では $v = +1$) のいずれかを選ぶ．根を選んだ場合は $t = 0$ からの差分を報告することになり，収集者は $\hat{\boldsymbol{f}}^0$ から直接

表 2 トイデータに対する 1-hot DDRM（多値拡張）の動作例

ユーザ（担当）		$t=1$	$t=2$	$t=3$
u_1 (A)	ビット d^t	1	1	0
	葉 c^t	+1	0	−1
	根	—	+1	—
u_2 (B)	ビット d^t	0	0	0
	葉 c^t	0	0	0
	根	—	0	—
u_3 (C)	ビット d^t	1	1	1
	葉 c^t	+1	0	0
	根	—	+1	—

$\hat{\boldsymbol{f}}^2$ を推定できるため，$t = 1$ の推定誤差が累積しない．$t = 3$ では u_1 の葉が -1（A の視聴終了）となる．チャンネル A の視聴率はグループ A（ここでは u_1 のみ）の校正済み報告の平均から推定され，B，C も同様である．u_1 は $t = 1, 2, 3$ で最大 3 回非ゼロ値を報告し得るため，$k < 3$ とした場合は 3 回目以降の非ゼロ値は 0 として（一様ランダムに）報告される．

4. 提案手法 Transition RR

4.1 アイデア

視聴データの性質（2)「視聴チャンネルは頻繁には変化しない」を利用するため，提案手法 Transition RR では，各時刻の状態そのものではなく，前時刻からの **状態遷移** (x_i^{t-1}, x_i^t) を RR の入力とする．状態に変化がない時刻には，入力に依存しない一様ランダムな遷移を出力する．このとき出力分布が入力に依存しないため，プライバシー予算を消費しない．予算を消費するのは状態が変化した時刻のみであり，各ユーザは高々 k 回の変化に予算 $\alpha = \varepsilon/k$ ずつを割り当てる．

視聴率は前時刻の推定値に遷移による変化分を加えて更新する．すなわち $\boldsymbol{f}^t = \boldsymbol{f}^{t-1} + \Delta\boldsymbol{f}^t$ であり，収集者は $\Delta\boldsymbol{f}^t$ を推定する．

変化を伴う遷移の集合を

$$\mathcal{T} = \{(a, b) \mid a, b \in \mathcal{X},\ a \neq b\} \quad (5)$$

とすると，その要素数は $L = M(M-1)$ である．遷移 $c = (a, b)$ が頻度ベクトルに与える変化を

$$\boldsymbol{\phi}(c) = \boldsymbol{e}_b - \boldsymbol{e}_a \quad (6)$$

と定義する．ここで $\boldsymbol{e}_j$ は第 j 成分のみが 1 の M 次元単位ベクトルである．$\mathcal{T}$ 上では正逆の遷移が対をなすため，$\sum_{c\in\mathcal{T}} \boldsymbol{\phi}(c) = \boldsymbol{0}$ が成り立つ．

4.2 摂動

摂動を加えるにあたり, c の属する集合に応じ摂動方式を変化させる．真の遷移が $c \in \mathcal{T}$ のとき，$\mathcal{T}$ 上の L-ary GRR

© 2026 Information Processing Society of Japan

$$\Pr[Y=y \mid c]=\begin{cases} p=\dfrac{e^{\alpha}}{e^{\alpha}+L-1}, & y=c,\\ q=\dfrac{1}{e^{\alpha}+L-1}, & y\neq c, \end{cases} \tag{7}$$

により出力 Y を得る．$c\notin\mathcal{T}$ のとき，または予算を使い切った後は，$\Pr[Y=y]=1/L$ で $\mathcal{T}$ から一様に出力する．

4.3 推定

$s=p-q=(e^{\alpha}-1)/(e^{\alpha}+L-1)$ とおく．真の遷移が c のとき

$$\mathbb{E}[\boldsymbol{\phi}(Y)\mid c]=p\,\boldsymbol{\phi}(c)+q\sum_{y\neq c}\boldsymbol{\phi}(y)=(p-q)\,\boldsymbol{\phi}(c)=s\,\boldsymbol{\phi}(c) \tag{8}$$

であり，変化がないときは $\mathbb{E}[\boldsymbol{\phi}(Y)]=\frac{1}{L}\sum_y\boldsymbol{\phi}(y)=\mathbf{0}$ である．したがって $\hat{\boldsymbol{\phi}}=\boldsymbol{\phi}(Y)/s$ はいずれの場合も不偏であり，収集者は

$$\Delta\hat{\boldsymbol{f}}^{t}=\frac{1}{n}\sum_{i=1}^{n}\frac{\boldsymbol{\phi}(Y_i^t)}{s},\qquad \hat{\boldsymbol{f}}^{t}=\hat{\boldsymbol{f}}^{t-1}+\Delta\hat{\boldsymbol{f}}^{t} \tag{9}$$

により頻度ベクトルを更新する．

4.4 プライバシー保証

定理 2 (Transition RR のプライバシー保証)**.** *Transition RR は T 時刻全体で ε-LDP を満たす.*

証明. ユーザの出力列を $\tilde{d}=(Y^1,\dots,Y^T)\in\mathcal{T}^T$ とし，任意の 2 つの入力系列 $d,d'\in\mathcal{X}^T$ を考える．時刻 t で予算を消費しない場合（変化なし，または予算を使い切った後）の出力は $\Pr[Y^t=y]=1/L$ であり入力に依存しないため，$\Pr[\mathcal{A}(d)=\tilde{d}]$ と $\Pr[\mathcal{A}(d')=\tilde{d}]$ の比において打ち消し合う．したがって比は予算を消費した時刻の項のみで定まる．各ユーザが予算を消費する報告を高々 k 回に制限し，1 回あたりの予算を $\alpha=\varepsilon/k$ とするとき，予算を消費する時刻は高々 k 回であるので, 式 (7) より出力確率は p または q のいずれかであるから，1 時刻あたりの比は高々 $p/q=e^{\alpha}$ である．ゆえに,

$$\frac{\Pr[\mathcal{A}(d)=\tilde{d}]}{\Pr[\mathcal{A}(d')=\tilde{d}]}\le\left(\frac{p}{q}\right)^{k}=(e^{\alpha})^{k}=e^{\varepsilon} \tag{10}$$

が成り立つ．よって定義 1 より ε-LDP を満たす． □

4.5 誤差上界

定理 3 (Transition RR の ℓ_2 誤差上界)**.** *各ユーザの状態変化回数が k 回以下であり予算枯渇が生じないとき，Transition RR の ℓ_2 誤差は*

$$\mathbb{E}\big[\ell_2^2\big]\le\frac{T(T+1)}{n}\left(\frac{1}{s^2}-\rho\right), \tag{11}$$

$$s=\frac{e^{\varepsilon/k}-1}{e^{\varepsilon/k}+M(M-1)-1}$$

表 3 トイデータに対する Transition RR の報告例（$k=2$）

ユーザ	$t=1$	$t=2$	$t=3$
u_1 真の遷移	OFF→A	—	A→B
報告例	OFF→A	B→C	A→B
予算	α	0	α
u_2 真の遷移	—	OFF→A	—
報告例	C→OFF	OFF→A	A→C
予算	0	α	0
u_3 真の遷移	OFF→C	—	—
報告例	OFF→B	A→OFF	B→A
予算	α	0	0

を満たす.

すなわち Transition RR の ℓ_2 誤差は T に対して線形のオーダーで増加する．

4.6 本問題への適用

トイデータでは状態が 4 値であるため $L=4\times3=12$ 通りの遷移がある．**表 3** に各ユーザの真の遷移と，報告例，予算消費の有無を示す．変化が生じるのは u_1 の $t=1,3$, u_2 の $t=2$，u_3 の $t=1$ の計 4 回であり，1 ユーザあたりの最大変化回数は 2 回である．$k=2$ とすれば各変化に $\alpha=\varepsilon/2$ が割り当てられる．変化のない時刻（表中の「—」）では一様ランダムな遷移が出力され，予算は消費されない.

たとえば $t=1$ における真の頻度変化は $\Delta\boldsymbol{f}^1=\frac{1}{3}\{(\boldsymbol{e}_{\mathrm{A}}-\boldsymbol{e}_{\mathrm{OFF}})+\mathbf{0}+(\boldsymbol{e}_{\mathrm{C}}-\boldsymbol{e}_{\mathrm{OFF}})\}$ であり $\boldsymbol{f}^1=(\frac{1}{3},0,\frac{1}{3},\frac{1}{3})$（A, B, C, OFF の順）が得られる．収集者は式 (9) により，報告された遷移を $\boldsymbol{\phi}$ で復号し $1/s$ 倍した平均を $\hat{\boldsymbol{f}}^{t-1}$ に加える．表の報告例では u_3 の $t=1$ の遷移が OFF → B に摂動されているが，$1/s$ による校正で期待値は真値に一致する.

5. 各手法のメリット・デメリット

Transition RR は定理 2 により, 全時刻全ユーザにおいて ε-LDP を満たす. また他 3 手法についても同様のプライバシーが保証される.

表 4 に 4 手法の特徴を整理する．GRR（時間分割）は実装が最も単純で相互排他性を定義域として利用できるが，予算が T 等分される．RR（時間・チャンネル分割）は相互排他性を利用せず，予算の分割数がさらに C 倍になるため最も不利である．1-hot DDRM は差分木により累積誤差を抑えつつ変化のない時刻の予算を節約できるが，多値拡張ではユーザを C 分割するため各チャンネルの標本数が n/C に減る．Transition RR は全ユーザを全チャンネルの推定に利用でき，遷移情報も保持されるが，報告ドメインが $M(M-1)$ と 2 乗で増加するためチャンネル数

© 2026 Information Processing Society of Japan

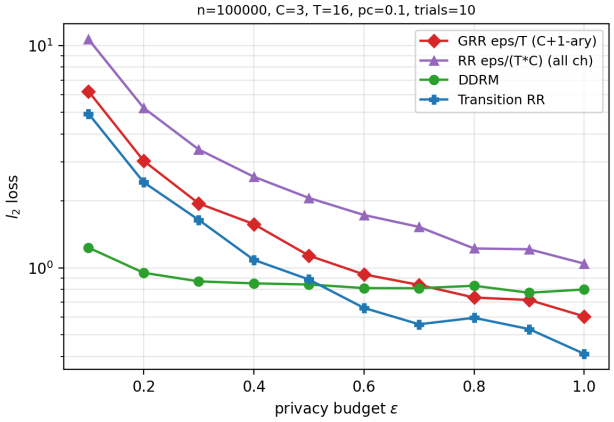

図 3 ℓ_2 誤差（$C = 3$，$\varepsilon \in [0.1, 1.0]$）

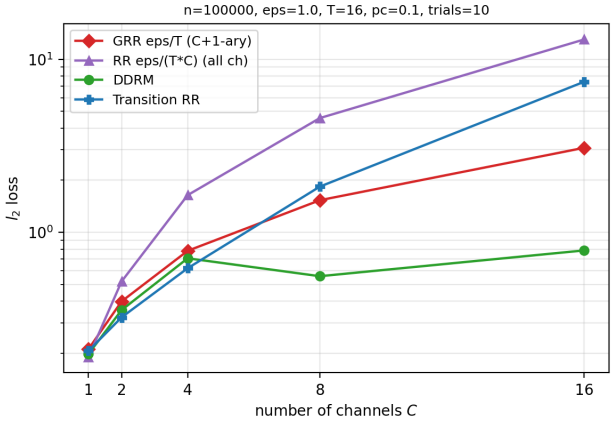

図 4 ℓ_2 誤差（$\varepsilon = 1.0$，$C \in \{1, 2, 4, 8, 16\}$）

が大きい場合に不利となる．なお本稿の Transition RR は連続的に差分を報告するものの，累積誤差についての処理は行っていないため，DDRM における差分木のような誤差補正が今後の課題として挙げられる．

6. 実験

6.1 実験設定

合成データを用いて 4 手法を比較する．各時刻で確率 p_c で現在と異なる状態へ一様ランダムに遷移し，確率 $1 - p_c$ で同じ状態に留まるようにデータを生成した．ユーザ数 $n = 100{,}000$ とし，変化しないパラメータは時間長 $T = 16$，チャンネル数 $C = 3$，変化率 $p_c = 0.1$ に固定する．1-hot DDRM, Transition RR で用いられるパラメータである予想変化回数 k は，DDRM における k の算出方法に従う [5]．機構の乱数を変えた 10 試行の ℓ_2 誤差の平均を報告する．1-hot DDRM および Transition RR の 予想変化回数 k は，DDRM[5] Theorem 7 の交点条件をそれぞれの機構のノイズ係数に置き換えて求めた値を用いる．

実験は次の 3 通りである．(i) 小さい予算 $\varepsilon \in [0.1, 1.0]$ で変化させる．（**図 3**），(ii) $\varepsilon = 1.0$ に固定し，チャンネル数 $C \in \{1, 2, 4, 8, 16\}$ で変化させる．（**図 4**），(iii) $\varepsilon = 1.0$ に固定し，時間長 $T \in \{2, 4, 8, 16, 32, 64\}$ で変化させる．（**図 5**）．なお ℓ_2 誤差は T 個の時刻にわたる誤差の二乗和の平方根であるため，(iii) では手法によらず T の増加とともに増大する．そこで (iii) では手法間の相対的な増加の速さに着目する．

また誤差の理論値と，実験的数値の比較として，n = 60,000, 変更率 $p_c = 0.2$ で合成データを用意し，プライバシー予算 ε= 1.0,k = 1 で，チャンネル数 C = 2, . . . , 10 まで変化させた時の，1-hot DDRM と Transition RR で頻度推定を行った．この実験では固定された予想変化回数 $k = 1$ を用いて実験を行った．

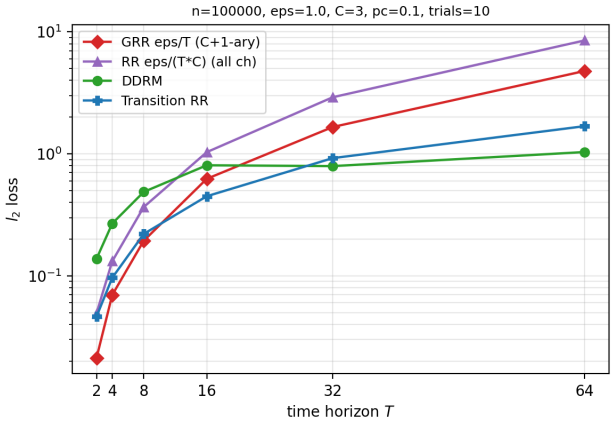

図 5 ℓ_2 誤差（$\varepsilon = 1.0$，$T \in \{2, 4, 8, 16, 32, 64\}$）

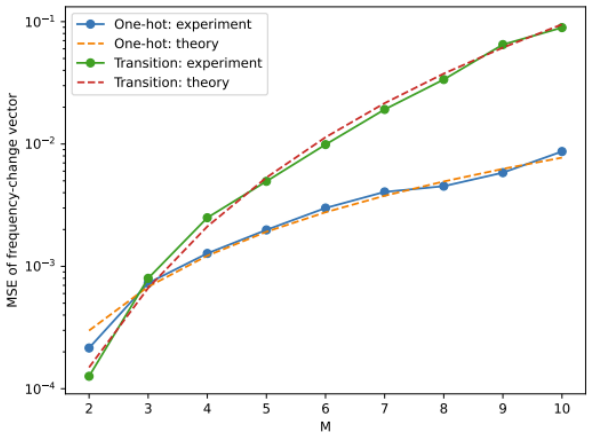

図 6 理論上と実験上の ℓ_2 誤差（$\varepsilon = 1.0$，$C \in \{2, 10\}$）

6.2 結果

6.2.1 予算 ε に対する比較

図 3 より，$\varepsilon \leq 0.5$ の小さい予算では 1-hot DDRM の誤差が約 0.8–1.3 で最も小さく，他手法は ε の減少とともに急激に悪化する．$\varepsilon \geq 0.6$ では Transition RR が最良となり，$\varepsilon = 1.0$ で約 0.5 に達する．GRR（時間分割）は Transition RR に次ぎ，RR（時間・チャンネル分割）は全

© 2026 Information Processing Society of Japan

表 4 4 手法の比較（$M = C + 1$）

	GRR（時間分割）	RR（時間・チャンネル分割）	1-hot DDRM	Transition RR
報告する情報	状態 x^t	one-hot 化した C ビット	担当ビットの差分	状態遷移 (x^{t-1}, x^t)
1 時刻あたりの報告回数	1 回	C 回（二値 RR を C 回）	1 回	1 回
報告する値域の大きさ	M	2	2（±1）	$M(M-1)$
1 報告あたりの予算	ε/T	$\varepsilon/(TC)$	ε/k（変化時のみ）	ε/k（変化時のみ）
無変化時の予算消費	あり	あり	なし	なし
各チャンネルの標本数	n	n	n/C	n
相互排他性の利用	あり	なし	なし	あり
累積誤差の補正	不要（各時刻独立）	不要	差分木	なし（葉のみ）
遷移情報の保持	なし	なし	なし	あり
ℓ_2 誤差上界	—	—	$2G_\alpha C\sqrt{T/n}$	$\sqrt{T(T+1)(1/s^2-\rho)/n}$
長所	単純・大予算で高精度	一般の二値データに適用可	C に対して頑健	小 C で高精度
短所	予算が T 分割	予算が TC 分割	標本数 n/C	ドメイン $O(M^2)$

域で最も悪い．1-hot DDRM の誤差は ε にほとんど依存せず約 0.8 で下げ止まる．これは，変化率が低いため予算消費が少なく ε の増加が 1 報告あたりのノイズをそれほど減らさない一方，ユーザ分割による標本数の減少と葉・根の 2 点のみによる累積誤差補正が誤差の下限を与えているためと考えられる．

6.2.2 チャンネル数 C に対する比較

図 4 より，$C = 1$ では 4 手法の誤差はほぼ等しい（約 0.2）．C の増加に伴い RR（時間・チャンネル分割）は $C = 16$ で約 13 まで，Transition RR は約 7.5 まで，GRR は約 3.1 まで悪化する．これに対し 1-hot DDRM の誤差は $C = 16$ でも約 0.8 に留まり，C に対して顕著に頑健である．また 1-hot DDRM の誤差が C の増加とともに単調増加ではない理由については、予算を使い切った後の完全ランダムな報告が影響していると考えられる．1-hot DDRM は各ユーザが 1 ビットのみを報告するため報告ドメインが C に依存せず，ユーザ分割による標本数の減少 (n/C) の影響が支配的となるためである．Transition RR は $C \leq 4$ では 1-hot DDRM とほぼ同等の最良の精度を示すが，報告ドメイン $M(M-1)$ が C の 2 乗で増加するため，$C \geq 8$ では GRR にも劣る．式 (**??**) の s の分母に $M(M-1)$ が現れることと整合する結果である．

6.2.3 時間長 T に対する比較

図 5 に $\varepsilon = 1.0$ における時間長 T の影響を示す．前述のとおり ℓ_2 誤差は T 個の時刻の誤差を累積するため，いずれの手法も T の増加とともに単調に増大するが，その増加の速さは手法間で大きく異なる．

T が小さい領域では，予算を分割する素朴な手法が有利である．これは，T が小さければ 1 報告あたりの予算 ε/T が十分大きく，予算分割の不利がほとんど生じない一方で，1-hot DDRM は T が小さいと差分木の階層が浅く，ユーザ分割による標本数の減少 (n/C) の不利のみが現れるためである．

T の増加に伴い順位は逆転する．$T = 16$ では Transition RR が最良となり，GRR，DDRM，RR の順となる．さらに $T = 32$ では 1-hot DDRM (0.80) と Transition RR (0.90) が拮抗し，$T = 64$ では 1-hot DDRM が約 1.0 で最良，Transition RR が約 1.8，GRR が約 5.0，RR が約 8.5 となる．$T = 16$ から $T = 64$ への 4 倍の延長に対する誤差の増加率は，1-hot DDRM が約 1.3 倍，Transition RR が約 4.0 倍であるのに対し，GRR は約 8.1 倍，RR は約 8.5 倍である．

また頻度推定の平均二乗誤差については, 誤差の理論値と実験で得られた誤差の結果がほぼ一致する結果となった. $k = 1$ で変化回数を固定したことにより, データ構造への依存が減少し, 実験で得られる誤差が理論上の誤差に近しくなったと考えられる．

7. 考察

まず理論的な考察として, 1-hot DDRM と Transition RR の誤差上界の比較を行う．式 (4) と式 (3) を比較すると，Transition RR の $O(T)$ に対し 1-hot DDRM は $O(\sqrt{T})$ であり，T が大きい領域で DDRM が優位となることが理論的に説明される．逆に 1 時刻あたりの係数は，1-hot DDRM が $2G_\alpha C$，Transition RR が $\sqrt{2}/s$ 程度であり，C が小さく T が中程度以下の領域では後者が小さくなり得る．ただし今回導出した誤差上界は, 十分な k を選択しユーザがその報告回数を超えない仮定のもとであるため, 予算を使い切った場合や十分な k の見通しが立たない場合の誤差上界については, さらなる実験による考察が必要である．

実験結果から，各手法が優位となる領域は次のように整理できる．チャンネル数が小さく ($C \leq 4$)，時間長が中程度以下 ($T \lesssim 16$) で，予算が中程度 ($0.5 \lesssim \varepsilon \lesssim 1$) の領域では Transition RR が最も高精度であり，全ユーザを利用できることと遷移情報を保持できることの利点が現れる．チャンネル数が大きい，時間長が長い ($T \gtrsim 32$)，または予算が非常に小さい領域では 1-hot DDRM が優位であり，

© 2026 Information Processing Society of Japan

差分木による累積誤差の補正と報告ドメインが C に依存しないことが効いている．予算が十分大きい，あるいは時間長が非常に短い領域では素朴な GRR が最良となる．

これらの結果は，各手法の不利がそれぞれ異なるパラメータに支配されることを示している．GRR と RR は予算分割数が T（あるいは TC）に比例するため T に，1-hot DDRM はユーザ分割による標本数 n/C の減少のため C に，Transition RR は報告ドメイン $M(M-1)$ のため C に，そして T にも影響を受ける．すなわち Transition RR は現状 C と T の双方に弱点をもち，これが $C \geq 8$ または $T \geq 32$ で他手法に劣る原因となっている．Transition RR のこの二つの弱点はいずれも機構の設計によって独立に緩和できる余地がある．

視聴率推定においては，地上波のようにチャンネル数が数個から十数個で，かつ強いプライバシー保証（小さい ε）が求められる状況が想定される．また視聴率は数分単位で長時間にわたり継続的に収集されるため，実用上は T が大きい領域が重要である．したがって，Transition RR の報告ドメインの小ささを保ちながら T に対する頑健性を 1-hot DDRM から取り込む，両者の中間的な設計が実用上重要である．

今後の課題として次を挙げる．第一に，Transition RR への累積誤差抑制機能の導入である．本稿の Transition RR が図 5 で $T \geq 32$ において 1-hot DDRM に逆転される主因は, 1-hot DDRM では補正されている累積誤差が顕出したと考えられる．例えば, 1-hot DDRM と同様に複数時刻にわたる遷移を上位ノードとして報告すれば，T に対する頑健性を得られると期待される．第二に，推定値の後処理である．視聴率は非負かつ総和が 1 以下という単体制約を満たすため，推定値をこの制約に射影することで分散を確実に減らせる．また番組表など公開情報を事前分布として用いた時間方向の平滑化も有効と考えられる．

8. おわりに

本稿では，視聴率の継続的推定を，相互排他性と低い変化率をもつ多値時系列データに対する LDP 頻度推定として定式化し，状態遷移を RR の入力とする Transition RR を提案し, 4 つの手法による比較実験を行った．時間分割型の GRR，時間・チャンネル分割型の RR，Transition RR, および 1-hot DDRM の多値拡張を同一の枠組みで比較し，合成データを用いた実験により，Transition RR はチャンネル数が小さく時間長が中程度以下で予算が中程度の領域において最も高精度であり，チャンネル数が大きい領域や時間長が長い領域では 1-hot DDRM が優位となることを示した．今後は Transition RR，DDRM をベースとした，C と T の双方に頑健で両者の利点を併せもつ機構の設計を進める．

参考文献

[1] Kasiviswanathan, S. P. *et al.*: What Can We Learn Privately?, *SIAM J. Comput.*, Vol. 40, No. 3, pp. 793–826 (2011).

[2] Ye, Q. and Hu, H.: Local Differential Privacy: Tools, Challenges, and Opportunities, *Proc. WISE*, pp. 13–23 (2020).

[3] Erlingsson, Ú., Pihur, V. and Korolova, A.: RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response, *Proc. CCS*, pp. 1054–1067 (2014).

[4] Ding, B., Kulkarni, J. and Yekhanin, S.: Collecting Telemetry Data Privately, *Proc. NeurIPS*, pp. 3574–3583 (2017).

[5] Xue, Q., Ye, Q., Hu, H., Zhu, Y. and Wang, J.: DDRM: A Continual Frequency Estimation Mechanism With Local Differential Privacy, *IEEE Trans. Knowl. Data Eng.*, Vol. 35, No. 7, pp. 6784–6797 (2023).

[6] Warner, S. L.: Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias, *J. Am. Stat. Assoc.*, Vol. 60, No. 309, pp. 63–69 (1965).

[7] Kairouz, P., Bonawitz, K. and Ramage, D.: Discrete Distribution Estimation under Local Privacy, *Proc. ICML*, pp. 2436–2444 (2016).

[8] Wang, T., Blocki, J., Li, N. and Jha, S.: Locally Differentially Private Protocols for Frequency Estimation, *Proc. USENIX Security*, pp. 729–745 (2017).

[9] Duchi, J. C., Jordan, M. I. and Wainwright, M. J.: Minimax Optimal Procedures for Locally Private Estimation, *J. Am. Stat. Assoc.*, Vol. 113, No. 521, pp. 182–201 (2018).

© 2026 Information Processing Society of Japan