情報セキュリティ (2016/7/5) レジュメ

1　データプライバシ保護

個人にとって秘密にしておきたい情報を含んだデータベースがあり，そのデータベースから統計情報などの有用な情報を外部に開示する状況を考える．このとき，プライバシを保護しながら情報を開示することはできるだろうか.

名前	性別	情報理論の成績	情報セキュリティの成績
青山 明男	男	S	B
井上 一郎	男	C	C
上野 歌代	女	S	C
江川 英太	男	A	B
音無 乙葉	女	S	S
片桐 和樹	男	B	不可
北野 喜一	男	B	C
工藤 国男	男	A	A
剣持 賢一	男	S	A
児玉 浩二	男	不可	不可

名前と性別は公開情報であり，成績は秘密情報だとする．このとき，以下の質問に答えることはできるだろうか.

1. 情報理論の単位を取得した学生は何人か.
2. 剣持健一は情報セキュリティの単位を取得したか.
3. 情報理論と情報セキュリティがともに不可の学生はいるか.
4. 情報理論で成績が S の女性は何人いるか.
5. 情報理論よりも情報セキュリティの成績の方がよかった学生は何人いるか.
6. 情報セキュリティの成績が A の学生は何人いるか.
7. 情報セキュリティの成績が A の学生は，工藤国男を除いて何人いるか.

病院等は医療情報として様々な統計情報を開示している．その他，収集した顧客データの統計情報を開示する企業もある．多くの場合，名前等の個人情報は隠されているが，予備知識，複数データベースの参照などによって個人情報が明らかになる可能性がある．また，個人情報が漏洩することを恐れて，情報提供者が真実でない情報を報告する可能性もある．そのような情報にもとづいた統計情報は有用なものとはいえない.

2　プライバシ保護のためのいくつかの技術

2.1　暗号化技術

プライバシを保護したいデータを暗号化し，第三者に内容がわからないようする技術である．暗号技術では，入力データを守ることはできるが，その内容をどのように外部へ開示すればよいかは考えない．

最近では，暗号化したままデータ処理を行う技術が発展しているが，どのような処理をすれば公開しても問題ないかという点は考えない．

2.2　匿名化

個人名等の秘密にしたい情報を明記せずに情報を収集する方法であり，アンケートへの回答でよく利用される．

複数の情報を照合することで，匿名化が破られる可能性がある (リンク攻撃)．攻撃者がどのような情報を持っており，また，将来的に持つことになるかを把握することは難しい．例えば，「試験のときに帽子をかぶっていた人は不合格だった」と匿名のまま情報を開示し，別の人が「なぜあの日ずっと帽子をかぶっていたの」と太郎に話しかければ，太郎の成績が明らかになる．

秘密にすべき情報が完全に特定されなかったとしても，ある程度候補が絞られてしまうとプライバシは保護されているとはいえないかもしれない．

2.3　k-匿名化

1998 年に提唱された考え方である．もとのデータベースに修正を加え（情報の一部を削除・一般化等），k-匿名性を満たすデータベースを作り，公開する．k-匿名性とは，データベースに存在するどのような属性の組を考えたとしても，それが一致する要素が k 回以上出現する状態のことである．これにより，どのようにデータを組み合わせても，k 人より少ない候補に絞ることはできない．

しかし，k-匿名性があっても，プライバシを保護していない状況が存在する．例えば，データベース x が与えられたときに，

$$M(x) = \begin{cases} \text{データを削除} & \text{井上一郎が情報理論の単位を取得したとき} \\ \text{データを一般化} & \text{上記以外のとき} \end{cases}$$

という操作を行っていた場合，k-匿名性は保たれるが，井上一郎の個人情報は保護されていない．

k-匿名化の難しい点は，データベースの内容に依存して操作を行う必要があることである．また，k-匿名化は，データベース全体を保護対象とするため，膨大なデータを扱いにくい．匿名化により，データベースから得られる統計情報が大きく劣化する可能性もある．

2.4 対話質問（差分プライバシ）

最初に示した，データベースとそれに対する質問の例を考えよう．質問の中には，単独でプライバシが守られていないことが分かるものもあるが，そうとも限らないものもある．6 と 7 の質問は，それぞれ単体で見るとプライバシが守られているように見えるが，この 2 つの質問を組み合わせると，工藤国男の秘密情報が明らかになってしまう．

このようなデータベースへの質問の回答を，プライバシを守りながら行う技術として**差分プライバシ** *(Differential Privacy)* と呼ばれる概念がある．提唱されてからまだ 10 年ほどだが，急速に発展している技術である．基本的なアイディアは，データベースへの質問（クエリ）に対して，ランダムな雑音を加えて回答するというものである．k-匿名化と比べると，差分プライバシでは，データベース自体への操作は必要ない．質問のあった情報を保護する技術であり，得られる統計情報の劣化を抑えることが期待できる．一方で，データベース自体は信頼できる所で管理する必要があり，データベースへの質問に対して管理者が適切に雑音を付加していることを信頼する必要がある．k-匿名化と異なり，公開された情報から正しく保護されていることを確認することが難しい．

以降では，差分プライバシの基礎を解説する．

3 プライバシを保護する対話質問

データベースプライバシでは，プライバシを保護しながら，データベースへの質問に対して有用な回答をする必要がある．

最初のデータベースの例において，質問に対して雑音を付加して回答する例を見てみよう．ここでは，質問 q に対し，正しい回答が A である場合に，$A+N$ という雑音を付加する方法を考える．（このような計算手順のことを**メカニズム**と呼ぶ．メカニズムとは単にアルゴリズムのことである．）雑音 N は，確率 30% で $N=0$，確率 18% で $N=\pm 1$, 確率 11% で $N=\pm 2$, 確率 6% で $N=\pm 3$・・・のように付加するものとする．

情報理論で成績が S の女性は何人いるかという質問へ回答することを考える．質問者は，情報理論で成績か S である確率は 40% であるという信念 (belief) をもっているとする．このとき，質問への回答を見ることで，信念がどの程度変化するかを考えてみよう．例えば，質問への回答が 1 であったとき，音無乙葉の情報理論の成績 S である確率はどのように変化するだろうか．イベント U, O はそれぞれ，上野歌代，音無乙葉の情報理論の成績が S である事象を表すものとする．求めたい確率は

$$\Pr[U \mid A+N=1] = \frac{\Pr[A+N=1 \mid U]\Pr[U]}{\Pr[A+N=1]}$$

である．

$$\begin{aligned}
&\Pr[A+N=1 \mid U,O] = \Pr[N=-1] = 0.18,\\
&\Pr[A+N=1 \mid \overline{U},O] = \Pr[N=0] = 0.3,\\
&\Pr[A+N=1 \mid U,\overline{O}] = \Pr[N=0] = 0.3,\\
&\Pr[A+N=1 \mid \overline{U},\overline{O}] = \Pr[N=1] = 0.18
\end{aligned}$$

であることから，

$$\begin{aligned}
\Pr[A+N=1 \mid U] &= \Pr[\overline{O}, N=0] + \Pr[O, N=1]\\
&= 0.6\cdot 0.3 + 0.4\cdot 0.18 = 0.252,\\
\Pr[A+N=1] &= \Pr[A+N=1 \mid U,O]\cdot\Pr[U,O] + \Pr[A+N=1 \mid \overline{U},O]\cdot\Pr[\overline{U},O]\\
&\quad + \Pr[A+N=1 \mid U,\overline{O}]\cdot\Pr[U,\overline{O}] + \Pr[A+N=1 \mid \overline{U},\overline{O}]\cdot\Pr[\overline{U},\overline{O}]\\
&= 0.18\cdot 0.4^2 + 0.3\cdot 0.6\cdot 0.4 + 0.3\cdot 0.4\cdot 0.6 + 0.18\cdot 0.6^2\\
&= 0.2376.
\end{aligned}$$

したがって，

$$\Pr[U \mid A+N=1] = \frac{0.252\cdot 0.4}{0.2376} \approx 0.42.$$

つまり，確率が 40% から 42% に変わったことが分かる．同様に，

$$\Pr[U \mid A+N=0] \approx 0.29, \Pr[U \mid A+N=2] \approx 0.41$$

と計算できる．質問への回答を知ることで，信念はそれほど大きくは変化していないことが分かる．

雑音を付加しない場合の確率と比較すれば，その効果が確認できる．

4 プライバシの定式化

データベース x とそれに対する質問 q が与えられたとき，回答 $M(x)$ を生成するメカニズム M を設計したい．メカニズム M に対して以下の性質を考える．

- **効用 (utility):** 回答 $M(x)$ は，x に q を質問したときの回答 $q(x)$ のよい近似である．
- **プライバシ (privacy):** 回答 $M(x)$ を見ても，x に関する信念は大きく変わらない．

効用は，$M(x)$ と $q(x)$ が近いことを要求している．どのように近さを表現すべきかは，必要とされる状況に大きく依存するため，ここでは深く考えない．例えば，標準偏差が分布の散らばり具合を表すものとすれば，標準偏差の逆数は分布のまとまり具合を表していると考えることができる．回答がよりまとまっている方が効用が高いと考えれば，

$$M \text{ の効用} = \frac{1}{\max_x \sqrt{E[(M(x)-q(x))^2]}}$$

とするのは 1 つの方法である．

プライバシでは，質問者の事前知識が，メカニズムの回答によって大きく変わらないことを要求している．しかし，質問者の事前知識が実際のデータベースの内容と大きく異なっていた場合，どのように回答しても信念が大きく変わってしまうことは避けられない．例えば，先程の例において，情報理論の成績が不可である確率は 90% であるという間違った事前知識をもっていた場合，よく近似された回答を受け取れば，信念は大きく変化してしまう．このような問題を回避するようにプライバシを定式化する必要がある．

上記の問題を回避するため，質問者はデータベースについて，特定の要素以外はすべて知っていると仮定する．そして，データベース x と，x から 1 要素だけ異なるデータベース x' を考え，各データベースから生成された回答は互いに「近い」ことを保証する．

■統計的距離 (変動距離) 分布 A と B の近さを表す概念として統計的距離があり，

$$\max_{T \subset Y} |\Pr[A \in T] - \Pr[B \in T]|$$

と定義される．ここで，Y は A および B がとりうる値の集合である．

しかし，以下の理由により統計的距離は近さの概念としてふさわしくない．データベース x の要素数は n とする．

- 距離 $\varepsilon \leq 1/(10n)$ だとする．この場合，要素を 1 つずつ変えることを考えると，$M(x)$ と $M(0^n)$ の距離は 0.1 以下となる．これは，回答 $M(x)$ は，確率 90% 以上で意味のないデータベース（0^n）からの回答であると考えることができる．
- 距離 $\varepsilon \geq 1/(10n)$ だとする．このとき，確率 1/10 でランダムに選んだ要素をそのまま出力するメカニズムを考えると，10 回に 1 回は確実に誰かの秘密情報を明らかにするため，プライバシを保護しているとは言えない．しかし，この操作自体は統計的距離の $1/(10n)$ 程度しか影響しないため，問題のない範囲と考えられてしまう．

■採用する分布間の距離 分布 A と B の間の距離が ε であることを，任意の $T \subset Y$ に対して，

$$\Pr[A \in T] \leq e^{\varepsilon} \Pr[B \in T]$$

を満たすときとする．近さを表す ε は $1/n \leq \varepsilon \leq 1$ の範囲を取ることを考える．（ε が小さすぎると，上記と同様の問題が生じる．）小さい ε に対して，$e^{\varepsilon} \approx 1 + \varepsilon$ が成り立つ．このことから，A と B の統計的距離は $O(\varepsilon)$ であることがわかる．ただし，単に統計的距離を抑えることよりも強いことを要求している．実際に，上記の 2 つ目の問題は，この定義では回避されている．こちらの定義では，$\Pr[A \in T]$ の値が小さい場合でも，$\Pr[B \in T]$ は，その e^{ε} 倍程度以下であることを要求している．統計的距離による定義では，発生確率が非常に小さい場合はその事象を無視しても大丈夫であったが，こちらの定義では無視できない．そのため，発生確率の小さい事象を無視せずに扱う場合に便利な定義であるといえる．

定義 1 (差分プライバシ)**.** メカニズム $M : \mathcal{D}^n \to \mathcal{Y}$ が ε-差分プライバシを満たすとは，*1* 要素だけ異なる任意のデータベースペア $x, x' \in \mathcal{D}^n$，取りうる任意の値 $y \in \mathcal{Y}$ に対して，

$$\Pr[M(x) = y] \leq e^{\varepsilon} \Pr[M(x') = y]$$

を満たすことである．

極端な場合として，$\varepsilon = 0$ を考えると，$\Pr[M(x) = y] \leq \Pr[M(x') = y]$ となり，x と x' を入れ替えた場合も成り立つことから，$\Pr[M(x) = y] = \Pr[M(x') = y]$ が成り立つ．これが任意の y で成り立つことは，分布 $M(x)$ と $M(x')$ が等価であることを意味し，それを満たすのは，M と x が独立な場合，つまり，$M(x)$ が全く有用でない場合に限られる．そのため，ε は 0 ではない小さい値を考える．

最初の計算例で見たような，信念の確率をもとに定式化することもできる．

定義 2. メカニズム $M : \mathcal{D}^n \to \mathcal{Y}$ が ε-ベイズプライバシを満たすとは，任意の $i \in \{1, \ldots, n\}$, $\mathcal{D}^n$ 上の任意の分布 X，任意の述語 $P : \mathcal{D} \to \{0, 1\}$，$M(X)$ が取りうる $y \in \mathcal{Y}$ に対し，

$$e^{-\varepsilon} \Pr[P(X_i)] \leq \Pr[P(X_i) \mid M(X) = y] \leq e^{\varepsilon} \Pr[P(X_i)]$$

を満たすことである．ここで，X_i は X の i 要素目を表す．

ベイズプライバシの方が，式の表している意味がわかりやすい．一方で，差分プライバシは，事前・事後などを気にする必要がなく，2 つのデータベース間の比較だけを行えばよく，式として扱いやすい．さらに，ベイズプライバシを示すには差分プライバシで十分であることが分かる．

定理 3. M が ε-差分プライバシを満たすとき， M は ε-ベイズプライバシを満たす．

5 ラプラスメカニズム

データベース $x \in \mathcal{D}^n$ に対し，述語 $P : \mathcal{D} \to \{T, F\}$ に関連付けられた**数え上げ質問** *(counting query)* q_P を以下のように定義する．

$$q_P(x) := (P(x_i) = T \text{ である要素 } i \text{ の数}).$$

数え上げ質問等の，数値を回答する質問に対するメカニズムとして，**ラプラスメカニズム**を紹介する．

定義 4. 尺度パラメータ b の（離散版）ラプラス分布 $\mathrm{Lap}(b)$ は，整数値 t の出現確率が

$$\Pr[N = t] = \frac{1}{Z} e^{-\varepsilon |t|}$$

で与えられる離散確率分布である．ここで，$Z = \sum_{t=-\infty}^{\infty} e^{-\varepsilon |t|}$ は正規化因子である．平均は 0，分散は $\sigma^2 = 2b^2$ である．

プライバシパラメータ $\varepsilon > 0$ のラプラスメカニズムは，数え上げ質問 q に対し，$M(x) = q(x) + N$ を答える．ここで，N はラプラス分布 $\mathrm{Lap}(1/\varepsilon)$ に従って選ばれる．

（通常，ラプラス分布は連続分布として定義され，連続版によってラプラスメカニズムを定義することもできる．連続分布版の方が一般性は高く，数値を回答する質問に広く応用できる．ただし，連続分布の場合は出力が実数のため，丸め込み操作が必要である．）

定理 5. ラプラスメカニズムは，数え上げ質問に対して ε-差分プライバシを満たす．

証明. 要素が 1 つだけ異なる $x, x' \in \mathcal{D}^n$ を考える．数え上げ質問 q に対しては, $|q(x)-q(x')| \leq 1$ が成り立つことに注意する．任意の $y \in \mathcal{Y}$ に対して，

$$\Pr[M(x) = y] = \Pr[q(x) + N = y] = \Pr[N = y - q(x)] = \frac{1}{Z} e^{-\varepsilon|y-q(x)|}$$
$$\leq \frac{1}{Z} e^{-\varepsilon|y-q(x')|+\varepsilon} = e^{\varepsilon} \cdot \frac{1}{Z} e^{-\varepsilon|y-q(x')|} = e^{\varepsilon} \Pr[M(x') = y].$$

したがって，ラプラスメカニズムは ε-差分プライバシを満たしている． □

このメカニズムの効用についても考えてみよう．標準偏差の逆数を効用だと考える．ラプラス分布 $\mathrm{Lap}(1/\varepsilon)$ の標準偏差 $\sigma = \sqrt{2}/\varepsilon$ であるため，効用は $\varepsilon/\sqrt{2}$ である．つまり，プライバシパラメータ ε を小さくすればするほど，効用が小さくなり，プライバシと効用にトレードオフの関係がある．

6 指数メカニズム

質問の回答が数値でない場合にも利用できるメカニズムとして，指数メカニズムがある．次の例を考えよう．

太郎，二郎，花子，明子の 4 人に，五郎島金時をせり（競売）で購入してもらうことを考える．付け値は，1 本あたり太郎が 300 円，その他の 3 人が 100 円とする．このとき，収入を最大化したいと考えているとする．

収入は，値段を 100 円にすると 4 人が購入するため 400 円，200 円にすると太郎だけが購入し 200 円となる．また，300 円にすると 300 円，400 円にすると 0 円になることがわかる．したがって，単に収入を最大化するには値段を 100 円にすればよい．

しかし，付け値を知らせることによって，太郎たちのプライバシが破られるかもしれない．例えば，太郎は金持ち（1 本 300 円までなら買う）で，二郎と花子は貧乏（1 本 100 円までなら買う）だったする．明子については何もわからないときに，収入を最大化する値段が明らかにされると，明子の予算が明らかになる．

指数メカニズムは，このような状況において，プライバシを守りながら収入を増やす値段をつけることを可能にする．上記の場合，付け値のデータベース $x \in \mathcal{D}^n$ を入力として受け取り，ある範囲 $\mathcal{R}$ の値段を出力するのが目的である．データベース x に対し，今回考えたい質問 q は，「五郎島金時の値段はいくら」という質問である．ここで，$x \in \mathcal{D}^n$ と $r \in \mathcal{R}$ に対して，効用スコア (utility score) $u(x, r)$ を定める．効用スコアは，最大化したい値の関数を表す．上記の例では，$\mathcal{R} = \{100, 200, 300, 400\}$ と値段の範囲を定め，値段 r のときの収入を $u(x, r)$ とすればよく，$u(x, 100) = 100$, $u(x, 200) = 200$, $u(x, 300) = 300$, $u(x, 400) = 0$ となる．

メカニズムの出力によってプライバシが守られない場合があるため，指数メカニズムでは出力を確率的に選ぶ．その確率は効用スコアに依存しており，スコアが高いものがより高い確率で選ばれるようにする．要素が 1 つだけ異なる $x, x' \in \mathcal{D}^n$ と取りうる $r \in \mathcal{R}$ をすべて考えたときの $|u(x, r) - u(x', r)|$ の最大値を Δu とおく．

■指数メカニズム　データベース x と効用スコア $u(x,r)$ がすべての $r \in \mathcal{R}$ について与えられたとき，r を $\exp(\varepsilon u(x,r)/2\Delta u)$ に比例した確率で出力する．

定理 6. 指数メカニズムは ε-差分プライバシを満たす．

指数メカニズムの効用について，次の定理が成り立つ．

定理 7. データベース $x \in \mathcal{D}^n$ に対し，$u^*(x) = \max_{r\in\mathcal{R}} u(x,r)$ とおく．このとき，任意の $t > 0$ に対し，指数メカニズムの出力の効用が，$u(x,r) - t$ 以下である確率は，$|\mathcal{R}| \exp(-\varepsilon t/2\Delta u)$ 以下である．