プライバシ強化技術（PETs）講義スライド

KDDI総合研究所　三本知明

講義の内容

プライバシに関する法規制について

・国内外のプライバシに関する法律の概要

・パーソナルデータに潜むリスク

プライバシ強化技術（１）：各プライバシ強化技術の基本

・匿名化

・差分プライバシ

・局所差分プライバシ

プライバシ強化技術（２）：各プライバシ強化技術の基本・応用例

・サンプリング

・匿名化アルゴリズム

・その他のプライバシ強化技術

・差分プライバシの応用例

・局所差分プライバシの応用例

プライバシに関する法制度について

データ利活用とプライバシ保護

データ利活用の広がり

・スマートフォンの普及やIoTなどの技術的進展により、あらゆる情報がデータ化され、利活用できるデータ量が増大

・加えてAI技術などによる蓄積データの分析が可能になりつつある

プライバシ保護の重要性

・データ利活用が進むことで、パーソナライズ化されたレコメンデーションなど生活の利便性が高まる

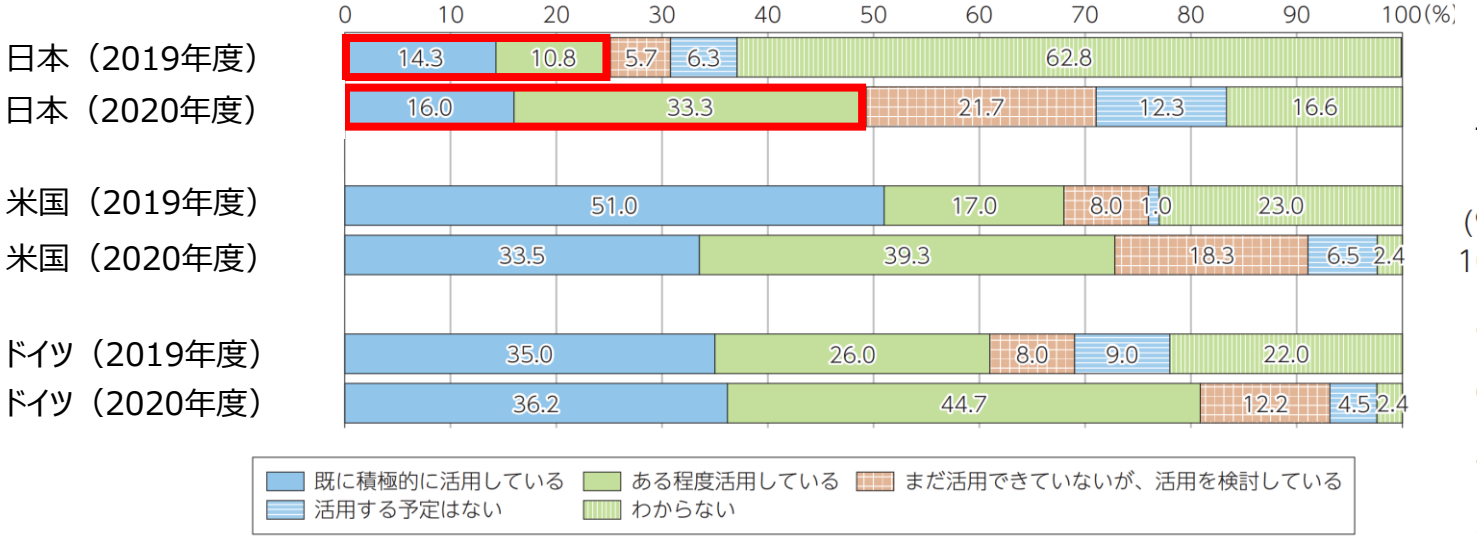
・その一方で個人に対するプライバシ侵害が危惧される

・プライバシ情報を利活用する際、データに携わるすべての関係者からのデータ漏洩の対策や法規則遵守が問題となる

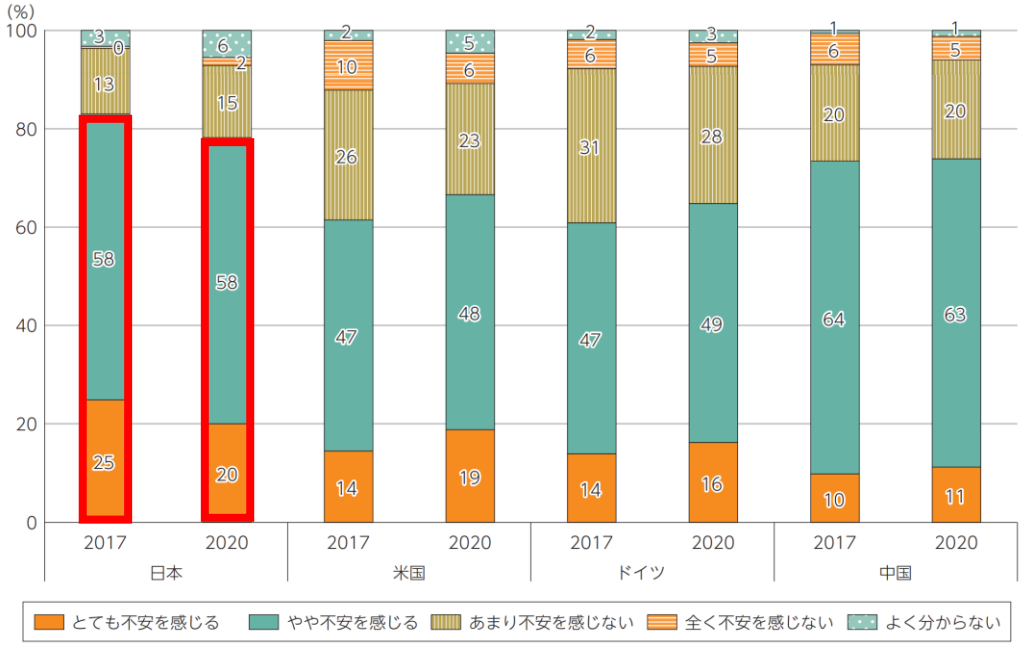
データ利活用とプライバシ保護

パーソナルデータの利活用状況

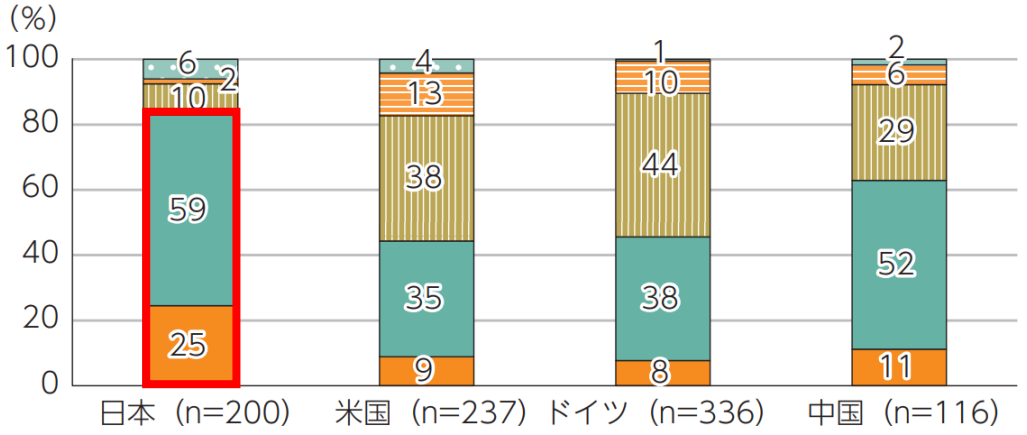
- 日本での利活用も拡大しているが、他国に遅れをとっている
- 多くの人がパーソナルデータを提供することに不安を抱いている
  - ー特に日本は他国と比べて不安感が強く、漠然とした不安を抱える人が多い



企業におけるパーソナルデータの活用状況



サービス・アプリケーション利用時にパーソナルデータを提供することについての不安



パーソナルデータを提供している認識がない人のうち、
パーソナルデータを提供することについての不安

参考：総務省情報通信白書
https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r03/pdf/n1200000.pdf
https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r02/pdf/n3300000.pdf

データ利活用とプライバシ保護

パーソナルデータ利活用に関する炎上事例

・2013年駅乗降履歴データの販売→**中止**

ー氏名、電話番号は除外していたが、細かい粒度で情報を提供していた

・2014年JR大阪駅ビルの監視カメラを用いた顔識別実験→**制限して再実験**

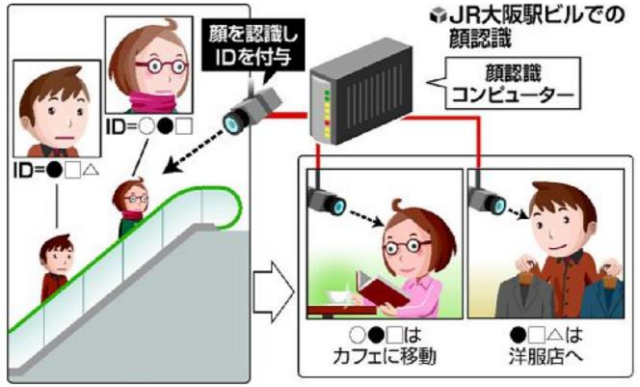
ー駅ビル内の人の行動を追跡し、避難誘導などの安全対策にむけた活用を予定していた

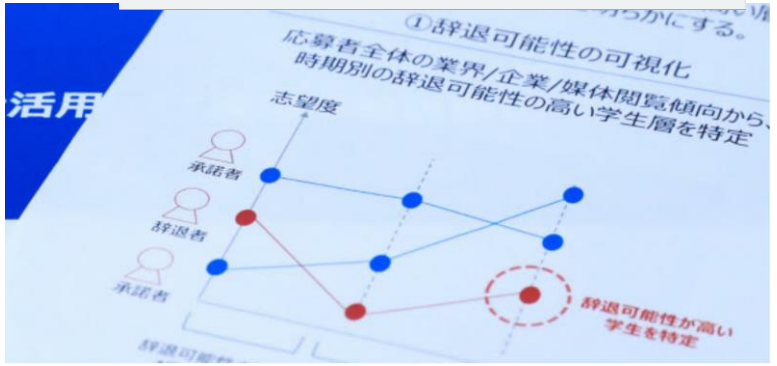
・2018年タクシーアプリの位置情報提供→**中止・データ削除**

ータクシー降車後に行った場所まで情報を提供していた

・2019年就活生の内定辞退率予測値の販売→**中止**

ー就活サイト上の行動履歴を分析して計測した内定辞退率を提供していた





炎上した原因

・事前説明が不足している

・データ活用の目的が不明瞭、公共性が低い

・個人に関する情報が得られてしまう危険性がある

参考URL
https://www.nikkei.com/article/DGXMZO48706050Z10C19A8TJ1000/
https://www.yomiuri.co.jp/net/news0/national/20140125-OYT1T00584.htm
https://www.nikkei.com/article/DGXMZO37182600R31C18A0X30000/?s=2

個人情報保護法の経緯

2003年 個人情報保護法成立（2005年全面施行）

・法施行後の10年間で、情報通信技術の発展

・制定当時には想定されなかったパーソナルデータの利活用が可能となる

2015年 個人情報保護法改正（2017年全面施行）

・３年ごと見直し規定が盛り込まれる

・国際的動向、情報通信技術の進展、新産業の創出・発展の状況等を勘案

2020年 3年ごと見直し規定に基づく初めての法改正

・個人の権利利益の保護

・技術革新の成果による保護と活用の強化

　ー越境データの流通増大に伴う新たなリスクへの対応

　ーAI・ビッグデータ時代への対応

・国際的な制度調和・連携

個人情報保護法の経緯

<u>個人情報保護法（2005年）における個人情報の定義</u>

・個人情報とは、生きている個人に関する情報であって、

ーその情報に書いてある名前や生年月日などにより特定の個人を識別できるもの

ー他の情報と簡単に照合することができて、それによって特定の個人を識別できるもの

個人情報保護法の経緯

改正個人情報保護法（2017年）

１．個人情報の定義の一部修正、**個人識別符号**の追加

・個人識別符号：その情報のみで特定の個人を識別できるもの

ー指紋、瞳の虹彩、パスポート番号、マイナンバー、ゲノム情報、etc.も個人情報の扱いに

２．機微な情報の取り扱いへの注意、**要配慮個人情報**の新設

・要配慮個人情報：本人の人権・信条・社会的身分・病歴・犯罪歴・犯罪被害の事実などが含まれる個人情報

・要配慮個人情報の取得は一部条件を除き同意取得が必須

ー本人の意思に優先すべき必要性が認められる場合（生命・身体・財産の保護が必要かつ同意取得が困難な場合 etc.）

ーすでに適正に公開されており、取得を制限する合理性がない場合（本人、あるいは報道機関による公開 etc.）

・要配慮個人情報の第三者提供には制限がある

ー個人情報を第三者提供する場合はオプトアウトによる外部提供が可能だが、要配慮個人情報は同意取得が必要

個人情報保護法の経緯

改正個人情報保護法（2017年）

３．オプトアウトの厳格化

・オプトアウト：事業者が個人情報を第三者提供しようとした場合に、本人からの反対がない限りこれに同意したものとみなし、個人情報を第三者に提供することができる仕組み

・オプトアウトには「本人が提供の停止を求めるのに必要な期間を置くこと」、「本人が所定の事項を確実に認識できる適切で合理的な方法によること」が必要

４．トレーサビリティの確保

・データ提供者は提供した日付、受領者名、住所等所定の事項を記録し、一定期間保存する必要がある

・データ受領者は受領した日付、提供者名、個人情報取得の経緯などを保存する必要がある

５．ペナルティ

・個人情報を取り扱う事業者がルール違反し、さらに国からの改善命令にも違反した場合

ー違反した従業員：最大６ヶ月の懲役または３０万円の罰金

ー会社：最大３０万円の罰金

・さらに民事でも被害者から損害賠償請求訴訟をされるリスクや、謝罪金支払いリスクがある

ーベネッセ個人情報流出事件の場合、損失額は200億円とも

個人情報保護法の経緯

改正個人情報保護法（2017年）

６．パーソナルデータの利活用促進、**匿名加工情報**の新設

・匿名加工情報：個人情報を、個人を識別できないように加工した情報

ー匿名加工情報は個人情報に当たらず、作成者と受領者はいくつかの義務があるものの、個人情報に関する取り扱いのルールは適用されない

・作成者の義務

ー適正加工義務：特定の個人情報であることを分からなくして、元になった個人情報を復元することが出来ないように加工する

ー安全管理措置：匿名加工情報の加工方法等の情報漏洩の防止や、苦情の処理などの管理措置を行う

ー公表義務：匿名加工情報に含まれている個人に関する情報の項目を公表する

ー第三者提供時の公表・明示義務：第三者提供する情報に含まれる個人に関する情報の項目と提供方法を公表する

ー識別行為の禁止義務：匿名加工情報を他の情報と突合して、個人を識別することの禁止

・受領者側の義務

ー安全管理措置：匿名加工情報の加工方法等の情報漏洩の防止や、苦情の処理などの管理措置を行う

ー第三者提供時の公表・明示義務：第三者提供する情報に含まれる個人に関する情報の項目と提供方法を公表する

ー識別行為の禁止義務：作成者の義務に加えて復元につながる情報を取得することも禁止

個人情報保護法の経緯

改正個人情報保護法見直し（2020年）

・個人の権利の拡充

・**仮名加工情報**の新設

　ー暗号化データなどの扱い

・厳罰化

1．個人の権利の在り方

① **利用停止・消去等の個人の請求権**について、一部の法違反の場合に加えて、**個人の権利又は正当な利益が害されるおそれがある場合等にも拡充**する。

② **保有個人データの開示方法**（現行では、原則、書面の交付）について、**電磁的記録の提供を含め、本人が指示できる**ようにする。

③ 個人データの授受に関する**第三者提供記録**について、**本人が開示請求できる**ようにする。

④ ６ヶ月以内に消去する**短期保存データ**について、保有個人データに含めることとし、**開示、利用停止等の対象**とする。

⑤ **オプトアウト**規定（※）により第三者に提供できる個人データの範囲を限定し、**①不正取得された個人データ、②オプトアウト規定により提供された個人データについても対象外**とする。

（※）本人の求めがあれば事後的に停止することを前提に、提供する個人データの項目等を公表等した上で、本人の同意なく第三者に個人データを提供できる制度。

令和4年4月以降に同規定による提供を行う場合は、令和3年10月１日より届出可能。

2．事業者の守るべき責務の在り方

① 漏えい等が発生し、個人の権利利益を害するおそれが大きい場合（※）に、**委員会への報告及び本人への通知を義務化**する。

（※）一定の類型（要配慮個人情報、不正アクセス、財産的被害）、一定数以上の個人データの漏えい等

② **違法又は不当な行為を助長する**等の**不適正な方法**により個人情報を利用してはならない旨を明確化する。

3．事業者による自主的な取組を促す仕組みの在り方

① 認定団体制度について、現行制度（※）に加え、**企業の特定分野(部門)を対象とする団体を認定できるようにする。**

（※）現行の認定団体は、対象事業者の全ての分野（部門）を対象とする。

4．データ利活用の在り方

① 氏名等を削除した**「仮名加工情報」を創設**し、内部分析に限定する等を条件に、**開示・利用停止請求への対応等の義務を緩和**する。

② 提供元では個人データに該当しないものの、**提供先において個人データとなることが想定される「個人関連情報」の第三者提供**について、**本人同意が得られていること等の確認を義務**付ける。

5．ペナルティの在り方　※令和2年12月12日より施行

① 委員会による命令違反・委員会に対する虚偽報告等の**法定刑を引き上げる**。

② 命令違反等の罰金について、法人と個人の資力格差等を勘案して、**法人に対しては行為者よりも罰金刑の最高額を引上げる**(法人重科)。

6．法の域外適用・越境移転の在り方

① 日本国内にある者に係る個人情報等を取り扱う外国事業者を、**罰則によって担保された報告徴収・命令の対象**とする。

② 外国にある第三者への個人データの提供時に、**移転先事業者における個人情報の取扱いに関する本人への情報提供の充実等**を求める。

※「７．その他」として、利用目的の特定、個人データの取扱いの委託及び公表等事項について、個人情報の保護に関する法律施行令等で規定

https://www.ppc.go.jp/files/pdf/r2kaiseihou.pdf

世界で広がるプライバシ保護規則

日本だけでなく世界中でプライバシに関する法規制の策定が進んでいる

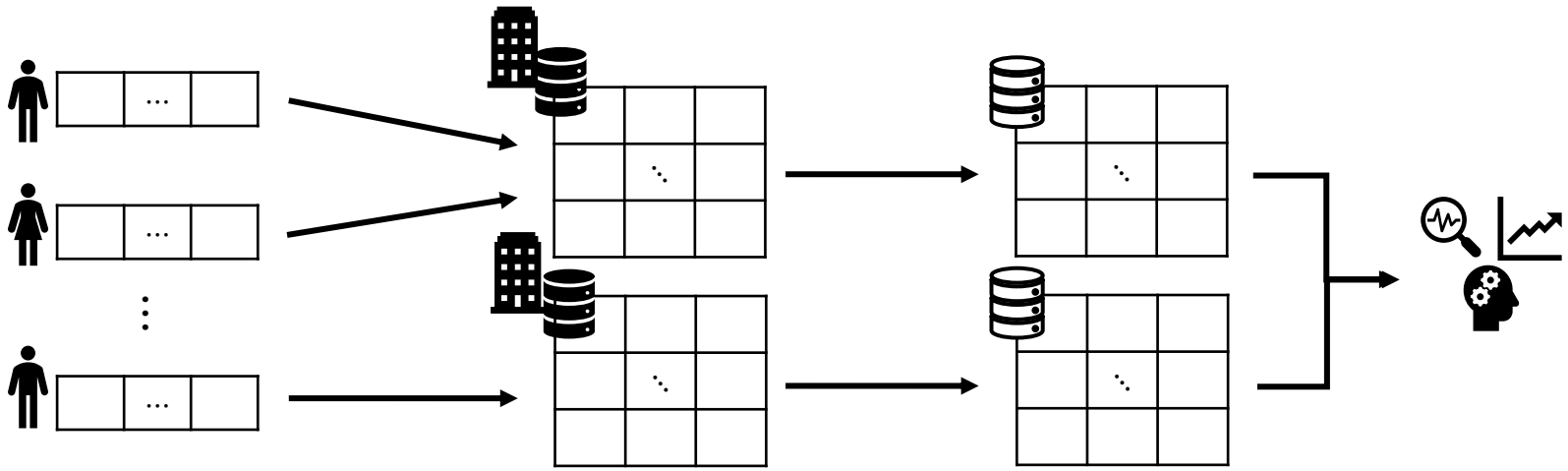
	2017	2020	2021	2023
日本	**改正個人情報保護法**	**個人情報保護法見直し**		
EU	**GDPR**			
米国		CCPA （カリフォルニア州）	VCDPA （バージニア州）	CPRA （カリフォルニア州）
ロシア			改正連邦法	
中国	サイバーセキュリティ法		データセキュリティ法草案 個人情報保護法草案	
韓国		個人情報保護法等データ３法改正		

プライバシ保護規則

GDPRなどのプライバシ保護規制は「ISO/IEC 29100 プライバシーフレームワーク」（2011）および「OECDプライバシー8規則」（1980）の考え方をベースとしている

ISO/IEC 29100	OECD	説明
同意及び選択 （Consent and choice）	収集制限（の原則） （Collection limitation principle）	データの収集、処理について明確な方法でデータ主体に同意を取ること
収集制限 （Collection limitation）		特定の目的のために、収集するデータを必要最低限に限定すること
目的の正当化及び明確化 （Purpose legitimacy and specification）	目的明確化 （Purpose specification principle）	データの収集前にその収集目的を明確化し、データの利用はその目的に限定すること
データの最小化 （Data minimization）		データにアクセスするユーザを最低限に抑え、知る必要性がない情報は得られないようにすること
利用、保持及び開示の制限 （Use, retention and disclosure limitation）	利用制限 （Use limitation principle）	データの利用、保持、提供を含む開示は、具体的、明示的かつ正当な目的を達成するために必要な範囲に限定すること
正確性及び品質 （Accuracy and quality）	データ内容 （Data quality principle）	収集したデータおよび処理されたデータが利用の目的に照らして、正確、完全、最新、十分かつ適切であること
公開性、透明性及び通知 （Openness, transparency and notice）	公開 （Openness principle）	データ収集の実施方法やポリシーを公開し、利用目的や利害関係者を明示、またデータ主体に対してデータの削除等を求める手段を開示すること
個人参加及びアクセス （Individual participation and access）	個人参加 （Individual participation principle）	データ主体が自分のデータにアクセス及び確認できるようにし、その正確性及び完全性に異議申し立てができ、データの修正や削除ができるようにすること
情報セキュリティ （Information security）	安全保護 （Security safeguards principle）	データの完全性、機密性および可用性を確保し、データのライフサイクル全体にわたって攻撃（権限外のアクセス、破棄、使用、変更、開示等）から保護すること
責任 （Accountability）	責任 （Accountability principle）	すべてのプライバシ関連のポリシー、手順および実践を適切に文書化し、共有すること
プライバシーコンプライアンス （Privacy compliance）		プライバシ保護に関する法令を遵守すること

データ活用の流れ



データ活用フロー	提供	加工	分析・活用	OECD8原則
本講義で触れる プライバシ強化技術	局所差分プライバシ			目的明確化 ・収集制限 ・利用制限
		匿名化・サンプリング		
		差分プライバシ		
		秘密計算		
		連合学習		
その他リスク	不正入手、改ざん、不正利用、漏洩			
その他対策	同意取得、暗号化、ブロックチェーン、認証、アクセス制御			個人参加・安全保護 ・データ内容・公開・責任

プライバシに関する法制度まとめ

プライバシに関する法規制が世界中で制定されている

・日本においては改正個人情報保護法が成立、３年おきに見直しがされる

　ー個人情報の定義、要配慮個人情報、匿名加工情報などの整理

・日本以外でも世界各国で法規制が進む

　ーGDPRは日本においても適用されるため、注意が必要

プライバシ強化技術は多種多様

・データ漏洩やトレーサビリティへの対応として暗号化技術やブロックチェーン技術が期待される

・目的明確化や収集制限、利用制限を目的としたプライバシ強化技術として、匿名化や差分プライバシなどが利用される

パーソナルデータにおけるプライバシリスク

パーソナルデータ提供に関するプライバシの懸念

１．個人特定

特定の個人と一意に結びつく情報が取り除かれ、どの個人に関するデータであるか分からないデータについて、そのデータを該当する個人と再び結びつけること

２．連結

ある個人に関するデータを同一人物に関する別のデータと結びつけること。特定の個人が特定されている・いないに関わらず起こりうる

３．属性推定

ある個人に関するデータの一部が削除、あるいは抽象化されているときに、それを復元、あるいは推定すること。特定の個人が特定されている・いないに関わらず起こりうる

４．濡れ衣

ある個人に関するデータが属性推定された際に、その属性に関して好ましくない推定を引き起こすこと。例えばある個人がある飲食店にいたと推測されたとする。さらにこの飲食店から感染症患者が発生した場合、その個人が感染者であると疑われうる

５．連絡

ある個人に関するデータを保持する者が、何らかの手段でその個人に連絡（訪問、郵便物を送る、電話をかける etc.）すること。特定がなくとも、無作為にメールを送るなどの連絡による被害が起こりうる。

６．直接被害

ある個人に関するデータを保持する者が、その個人に直接的な被害を与える。例えばその個人のクレジットカード番号を無断使用する、SNSのアカウント情報を使用して無断で書き込む。特定がなくとも、無作為にクレジットカード番号を無断利用するなどの直接被害が起こりうる。

プライバシ侵害の事例

マサチューセッツ州知事特定の事例

・マサチューセッツ州のGroup Insurance Commition (GIC)では13.5万人の州職員とその家族の医療保険に関する情報（氏名、性別、郵便番号、生年月日、治療内容、請求総額等）を収集していた

・GICは氏名を取り除いたデータを民間企業に販売していた

・一方でマサチューセッツ州ケンブリッジの選挙人名簿は購入することができ、これには氏名、性別、郵便番号、生年月日、支持政党などが含まれていた

プライバシ侵害の方法

・連結による個人特定が可能

ー医療保険データが選挙人名簿との照合によって個人が特定できる情報に復元され、州知事のレコードが特定された
（州知事と同じ生年月日が6人おり、うち3人が男性、郵便番号が同じ人は他にいなかった）

・連結による属性推定が可能

ー州知事の医療保険に関する情報が容易に推定可能

プライバシ侵害の事例

Netflixの事例

・Netflixは推薦アルゴリズムのコンペを目的として、全データのうち10%以下にサンプリングされ、利用者を直接的に特定する情報を削除した上で約48万人の映画のレイティング値約1億件を提供した

プライバシ侵害の方法

・攻撃者が以下のような背景知識を持つ場合、個人の特定が可能

ーその個人が過去に与えた8つの映画のレイティング値を知っており、そのレイティング値を与えた日付が2週間単位の精度であれば、99%の確率でその個人のレコードを特定可能

ーその個人が過去に与えた2つの映画のレイティング値を知っており、そのレイティング値を与えた日付が3日間単位の精度であれば、68%の確率でその個人のレコードを特定可能

・異なるデータベースとの突合による連結が可能

ー公開されているInternet Movide Database (IMDb)から取得できるデータとレイティング値をもとに連結が可能であり、IMDbに含まれるレビューコメントから趣向や政治的傾向等が推定可能

（50人のIMDbユーザのうち、2人のデータがNetflixのユーザであることが高い確度で推定され、ユーザが残した映画のレビューには強い政治的傾向が見られた）

プライバシ侵害の事例

Taxi Rideの事例

・New York City Taxi and Limousine Commissionは法律（FOIL）にもとづいて約1.7億レコードのタクシー乗降履歴（乗車地点、降車地点、乗車時間、料金等）を提供していた

・また個別の乗車記録がどのタクシーの記録であるかを特定されないようにタクシー運転手の免許番号とタクシーのナンバーから仮IDを付与していた

プライバシ侵害の方法

・知識があれば仮IDから元の免許番号やナンバーに戻すことが可能

ー同一タクシーには同一の識別番号が付与されていたため、各タクシードライバーの移動履歴や料金から、
それぞれのドライバーの住所や収入が推測可能

ーナンバーからタクシーの特定が可能であり、乗客の行き先が特定可能
（ある俳優がGreenwich Villageで降り、10.5$支払ったことが特定された）

・ある地点に降車した客がどこから乗車したかの推定が可能

ーセンシティブな場所（賭博場や感染症専門の病院等）に降車したユーザの自宅や職場の推測が可能

パーソナルデータの構成要素

属性情報の種別

・識別情報

・履歴情報

・要配慮情報

・連絡情報

識別情報

・**直接識別情報（識別子）**：それ単体で個人の特定を可能にする情報

（e.g.）氏名、指紋データ、遺伝子情報、マイナンバー、運転免許番号 etc.

・**間接識別情報（準識別子）**：個人に関する不変的な情報で、複数組み合わせることで個人を識別しうる情報

（e.g.）生年月日、性別、郵便番号 etc.

パーソナルデータの構成要素

履歴情報

・個人の活動にかかわる履歴の情報を履歴情報と呼ぶ

（e.g.）移動履歴、購買履歴、Web検索履歴

・履歴情報は識別情報として働く場合がある

ー特異性：履歴情報の一部に特異な値が含まれる場合

（e.g.）ほとんど売れない高額商品

ー習慣性：特定の値が頻出し、習慣的な不変の属性値が推測される場合

（e.g.）住宅から会社までの移動

ー一意性：同一の履歴情報が存在しない場合

（e.g.）10年間蓄積された履歴情報

「トヨタ、顧客情報215万人分が10年近く閲覧可能な状態に」
https://www.nikkei.com/article/DGXZQOFD124YP0S3A510C2000000/

パーソナルデータの構成要素

要配慮情報

・個人の特定に結びつかなくとも、単体で取り扱いに配慮が必要な情報

（e.g.）人種、国籍、宗教、犯罪歴、病歴 etc.

連絡情報

・個人への連絡を可能とする情報

（e.g.）メールアドレス、電話番号、会社名、住所 etc.

各情報の境界は非常にあいまい

・あるデータが直接識別情報や間接識別情報、要配慮情報となるかは文脈によって変わる

ー検索履歴に宗教に関する単語が含まれている場合や移動履歴に感染症専門の病院が含まれている場合、
これらの履歴情報は要配慮情報にもなり得る

ー母集団が30人のクラス名簿において、生年月日は直接識別情報になり得る

ーメールアドレスに名前が含まれている場合は直接識別情報になり得る

・パーソナルデータを扱う際は、各情報がどのような特性を持つかを十分に議論する必要がある

データ提供のリスク

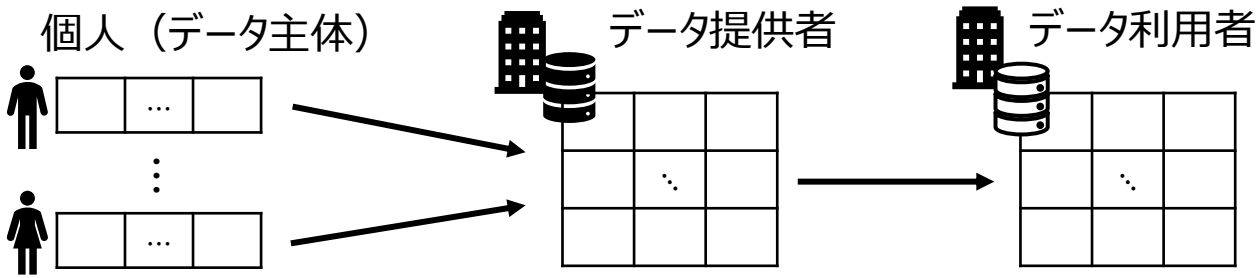
一般にプライバシ情報を第三者から守りながらデータ活用をしたい場合、データの解析結果のみを提供した方がプライバシの観点から問題が少ない

データ提供が必要となる状況

1．データ利用者の解析目的がデータ取得時点で決まっていない
2．データ提供者がデータ解析を実現する技術や計算資源を持たない
3．データ利用者の持つ他のデータを組み合わせてデータ解析を行う
4．データ利用者の解析目的や解析手法を第三者に明かしたくない

データ提供のプロセス

1．データ提供者は個人からデータを収集する
2．データ提供者は、収集したデータを加工し、データ利用者に提供する
3．データ利用者は、提供されたデータを用いてデータ解析を行う



データ提供のリスク

データ提供のリスク

・電話番号やクレジットカード番号、マイナンバー、アカウント情報など個人に直接紐づく情報は連絡、直接被害に繋がる自明なリスク

・連結や属性推定などは直接識別子情報がなくとも起こりうる非自明なリスク

　ーデータ提供の方法や攻撃者が持つ知識、データの分布、加工方法など様々な要因によってリスクが変化する

攻撃者モデルの仮定

・プライバシ保護：データ提供者は提供するデータDから個人が特定されないようにデータ加工を行い、D'を生成する

・攻撃：攻撃者はデータ利用者としてD'を入手し、外部情報（背景知識）を利用してD'の個人を連結、あるいは特定する

データ提供のリスク

<u>データ提供における想定される攻撃：医療保険会社が保有するデータを物販会社に提供するケース</u>

・マイナンバー、氏名、住所は直接識別情報に類する情報

・年齢、性別、職業は間接識別情報に類する情報

・その他はいずれにも該当しない

マイナンバー	氏名	年齢	性別	住所	職業	血圧	…	既往歴
xxx1	浅井	35	男	東京都A区B丁目C番	医師	135mmHg	…	肺がん
yyy2	井上	28	女	大阪府Y市Z町○○-○	公務員	142mmHg	…	糖尿病
zzz3	上田	39	男	京都府P市Q町X-X-X	会社員	118mmHg	…	胃潰瘍
…	…	…	…	…	…	…	…	…

医療保険会社によるデータ加工

仮ID	年齢	性別	住所	職業	血圧	…	既往歴
A052	35	男	東京都	医師	135mmHg	…	肺がん
27C5	28	女	大阪府	公務員	142mmHg	…	糖尿病
81D5	39	男	京都府	会社員	118mmHg	…	胃潰瘍
…	…	…	…	…	…	…	…

データ提供のリスク

<u>データ提供における想定される攻撃：医療保険会社が保有するデータを物販会社に提供するケース</u>

・直接識別情報がなくとも物販会社による連結が起こる

・物販会社は既往歴や職業に応じて、個人に広告を送るなどができる

ー個人情報保護法ではこのような特定を行うことは禁じられている

ID	年齢	性別	住所	職業	血圧	…	既往歴
A052	35	男	東京都	医師	135mmHg	…	肺がん
27C5	28	女	大阪府	公務員	142mmHg	…	糖尿病
81D5	39	男	京都府	会社員	118mmHg	…	胃潰瘍
…	…	…	…	…	…	…	…

医療保険会社が提供するデータ

マイナンバー	氏名	年齢	性別	住所	…	メールアドレス	ユーザID
www4	遠藤	41	男	兵庫県	…	endo@mail.com	Hyoo
yyy2	井上	28	女	大阪府	…	ino@mail.com	B_1994
vvv5	小川	18	男	埼玉県	…	oga@mail.com	Saitama18
…	…	…	…	…	…	…	…

物販会社が保有するデータ

日時	ユーザID	製品番号	製品価格
23/5/1 19:03	Hyoo	P001	150円
23/5/1 19:03	Hyoo	P028	280円
23/5/1 19:07	XXX	P005	170円
…	…	…	…

パーソナルデータにおけるプライバシリスクまとめ

パーソナルデータを構成する要素

・パーソナルデータには直接識別情報、間接識別情報、履歴情報、要配慮情報などがあるが、明確に分類することは難しい

パーソナルデータを利用した攻撃

・プライバシ侵害として個人や属性の特定から連絡や直接被害、濡れ衣などが起こる

・直接識別情報を削除するだけではプライバシリスクを十分に抑えることができない

匿名化

プライバシ強化のための主な手法

プライバシ強化を目的としたデータ加工の手法は以下に大別される

手法	具体的な手法	概要	具体例
一般化 (Generalization)	一般化	主に階層木に従って情報を一般化する	48歳 → 40代 大阪 → 関西
	トップ・ボトムコーディング	一定以上、以下の属性値を一般化する	110歳 → 80歳以上
抑制 (Suppression)	属性値削除	外れ値となる属性値を削除する	（170cm, 120kg） →（170cm, - kg）
	レコード削除	外れ値となるレコードを削除する	（150cm, 100kg） →削除
摂動 (Permutation)	ノイズ付与	実際の値に特定のノイズを付与する	14歳 → 21歳 サッカー → 水泳
	データスワッピング	レコード間の属性値を入れ替える	
サンプリング (Sampling)	ランダムサンプリング	データセットの一部だけを利用する	
仮名化 (Pseudonymization)	仮名の変更	同一人物のデータに対して、異なる疑似IDを付与する	
	暗号化 （秘密計算）	各値を全く異なる値に置き換える。追加情報（秘密鍵）があれば復元可能	

データ提供におけるリスク評価指標：k-匿名性

パーソナルデータを提供する際、直接識別情報を削除するだけでは特定や連結は防げない

・加工したパーソナルデータがどの程度プライバシ侵害のリスクがあるかを定量評価する必要がある

代表的なプライバシリスク評価指標：k-匿名性

・n人の個人からd個の属性を収集することを想定する

・個人 i から収集したデータを$\boldsymbol{x}_i = (x_{i1}, x_{i2}, \cdots, x_{id})$とし、全員分のデータ集合を$D = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$とする

・ d_{QI}個の属性$X_1, \cdots, X_{d_{QI}}$を間接識別情報とする

　−k-匿名性における攻撃者モデルでは、攻撃者は間接識別情報を背景知識として持つと想定する

・個人 i のデータを間接識別情報の組み合わせとその他の情報の組み合わせとして$\boldsymbol{x}_i = \left(\boldsymbol{x}_i^{QI}, \boldsymbol{x}_i^{other}\right)$とする

・このときk-匿名性は以下のように定義される

定義：k-匿名性

Dを n 人の個人から集めたレコードの集合とする。またDに含まれる間接識別情報の値の組み合わせをAとする。

すべての$x^{QI} \in A$について、x^{QI}を含むレコードがDに含まれない、あるいは少なくともk個存在するとき、Dはk-匿名性を有する。

データ提供におけるリスク評価指標：k-匿名性

k-匿名性を持つデータの具体例

直接識別情報　　間接識別情報

名前	年齢	性別	住所	年収	既往歴
青木	33	M	東京	1,100	うつ病
井上	34	M	埼玉	900	うつ病
	⋮	…	…	⋮	⋮
山田	47	M	京都	200	風邪
湯川	49	F	奈良	300	糖尿病
吉田	88	M	沖縄	10,000	肺がん

2-匿名化

10歳区切りで丸める（一般化）

地域単位で丸める（一般化）

年齢	性別	住所	年収	既往歴
30	M	関東	1,000	うつ病
30	M	関東	1,000	うつ病
⋮	…	…	⋮	⋮
50	-	近畿	250	風邪
50	-	近畿	250	糖尿病
88	-	沖縄	10,000	肺がん

外れ値の削除（レコード抑制）

属性値の削除（抑制）

グループ毎に平均化（マイクロアグリゲーション）

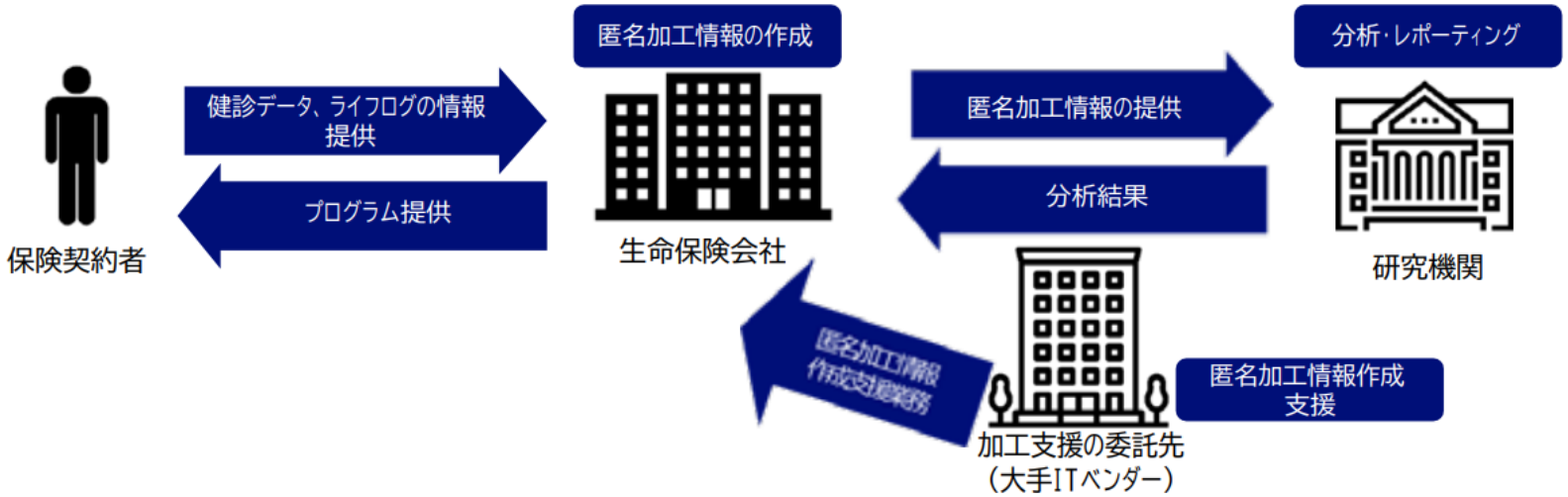
k-匿名性を持つデータの実用例

匿名加工情報の作成にはk-匿名性が意識されている

・生命保険会社による健康データ等の利活用事例

・その他電力（HEMS）データや位置情報データへの利活用事例
などが紹介されている

・データの匿名化方法以外に、安全管理措置や第三者提供時の契約に関する情報なども紹介されている



利活用全体像

項目	加工方法
属性データ	
年齢・生年月日	k-匿名化をベースとした加工（同じ属性の組合せを持つ個人が複数以上になるように加工）
性別	k-匿名化をベースとした加工（同じ属性の組合せを持つ個人が複数以上になるように加工）
住所	市町村単位までに加工（一般化）。また母数が少なく匿名性が確保できない地域は削除
職種	加工前のデータがいくつかの区分に分類されたデータであるため、該当者数が少ない職種を「その他」に加工している
年収	加工前のデータがいくつかの区分に分類されたデータであるため、加工なし
顧客 ID	ハッシュ関数による変換をして、別 ID に置換え
保険契約に関する情報	
保険種類	加工なし
払込保険料	千円単位に置換。またトップコーディング、ボトムコーディングにより外れ値を処理
健診データ	
検査値	外見（身長、体重等）に関する情報は、外れ値を処理（トップコーディング、ボトムコーディング）
健康増進プログラムに関する情報	
ライフログ（歩数、運動時心拍数等）	加工なし
ポイント獲得状況	加工なし

データ加工の実施方法と工夫点

個人情報保護委員会「匿名加工情報制度について」
https://www.ppc.go.jp/personalinfo/tokumeikakouInfo/

データ提供におけるリスク評価指標：属性推定に関する指標

属性推定リスク

・特定が起こっていないにも関わらず、個人の属性値が攻撃者に知られる可能性がある

・以下の例では2-匿名性を持つため、攻撃者は個人の特定はできないが、要配慮情報である既往歴が分かる
　ー青木、井上の特定は出来ないが、既往歴はうつ病で確定

濡れ衣リスク

・属性推定により不利益を被る可能性がある

・以下の例では山田の既往歴としてエイズを疑われる可能性がある

直接識別情報　　間接識別情報

名前	年齢	性別	住所	年収	既往歴
青木	33	M	東京	1,000	うつ病
井上	34	M	埼玉	1,000	うつ病
	⋮	…	…	⋮	⋮
山田	47	F	京都	200	風邪
湯川	49	M	奈良	300	エイズ
吉田	88	M	沖縄	10,000	肺がん

2-匿名化

年齢	性別	住所	年収	既往歴
30	M	関東	1,000	うつ病
30	M	関東	1,000	うつ病
⋮	…	…	⋮	⋮
50	-	近畿	250	風邪
50	-	近畿	250	エイズ

データ提供におけるリスク評価指標：属性推定に関する指標

定義：l-多様性

Dを n 人の個人から集めたk-匿名性を持つレコードの集合とする。またDに含まれる間接識別情報の値の組み合わせをAとする。すべての$x^{QI} \in A$について、x^{QI}を含むレコードの要配慮属性の種類が少なくとも l 個存在するとき、Dはl-多様性を有する。

その他属性推定リスクに対する評価指標

・エントロピーl-多様性：各x^{QI}を含むレコードの要配慮属性のエントロピーが$\log(l)$以上である時

・(c, l)-多様性：頻出する要配慮属性と希少な要配慮属性の出現頻度の差が以下の式で制御される時

$$r_1 < c(r_l + r_{l+1} + \cdots + r_m)$$

なお、r_iはx^{QI}において i 番目に出現頻度の高い要配慮情報の出現回数

・t-近似性：既知の要配慮属性の分布と各x^{QI}を含むレコードの要配慮属性の分布の距離がt未満である時

・その他： (X, Y)-プライバシ、(ϵ, m)-匿名性、FF-匿名性、etc.

年齢	性別	住所	年収	既往歴
30	M	関東	1,000	うつ病
30	M	関東	1,000	うつ病
⋮	…	…	⋮	⋮
50	-	近畿	250	胃がん
50	-	近畿	250	エイズ

2-多様化

年齢	性別	住所	年収	既往歴
30	M	関東	1,000	うつ病
30	M	関東	1,000	胃がん
⋮	…	…	⋮	⋮
50	-	近畿	250	うつ病
50	-	近畿	250	-

属性値の入れ替え（スワッピング）

属性値の削除（抑制）

データ匿名化手法

データの匿名化プロセス

１．直接識別情報を仮名IDに置き換える（**仮名化**）

２．間接識別情報を加工し、特定性を低減させる（**匿名化**）

ー要配慮情報や特異値に対する処理を含む

データ匿名化手法：仮名化

仮名化

・仮名IDは直接識別情報の代わりに個人を識別するために利用される情報

ー複数の個人に対して同一の仮名IDは割り当てられない

ーデータ提供者は直接識別情報と仮名IDを容易に紐付けられる

ーデータ利用者は直接識別情報と仮名IDを容易に紐付けられない

一方向性ハッシュ関数による仮名化

・一方向性ハッシュ関数とは任意の入力xについて、$H(x)$を求めることは容易だが、$H(x) = H(y)$となるx, yを求めることが困難であるような関数

・直接識別情報をxとして、ハッシュ値$H(x)$を仮名IDにする

ブルートフォースアタックによる仮名IDに対する攻撃

・ xのパターン数が数千万件程度であれば、現実的な時間で総当りによるxと$H(x)$の対応表の生成が可能

ーTaxi Rideの事例では、タクシーのナンバープレート（高々6桁の英字と数字の組み合わせ）をハッシュ関数の入力として利用

鍵付きハッシュ関数（SHA-1, MD5, etc.）による仮名化

・ブルートフォースアタックへの対応として、ランダムなビット列kを鍵とした**鍵付きハッシュ関数**の利用が適切

ー$H(x||k)$を仮名IDとして利用する

データ匿名化手法：匿名化

匿名化

・個人の特定リスクを低減するために間接識別情報を加工する操作を匿名化とよぶ

　ーデータセットがk-匿名性を満たすように加工することをk-匿名化とよぶ

・匿名化の手法には以下のようなものが存在する

　ー一般化

　ートップ（ボトム）コーディング

　ー抑制

　ーマイクロアグリゲーション

データ匿名化手法：匿名化

一般化（再符号化）

・一般化階層構造にもとづいてデータを加工することで特定リスクを低減する

　ー大域的再符号化：データセット全体に対して等しく一般化を行う

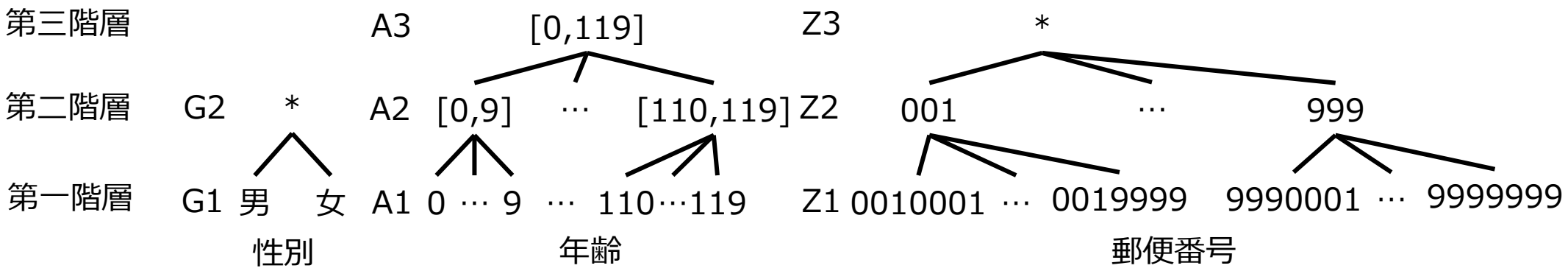
　ー局所的再符号化：任意のレコード群に対して局所的に一般化を行う

・属性の一般化階層構造はあらかじめ定義する必要がある

トップ（ボトム）コーディング

・ある閾値よりも大きい、あるいは小さいデータをまとめて一般化する

　ー例えば年齢が110歳など頻度分布の裾にあたる属性値は、それだけで特定リスクが高まるため[80,-]のように一般化する



データ匿名化手法：匿名化

抑制

・値を削除することで特定リスクを低減する

ーレコード抑制：k-匿名性を満たさないレコードを削除する

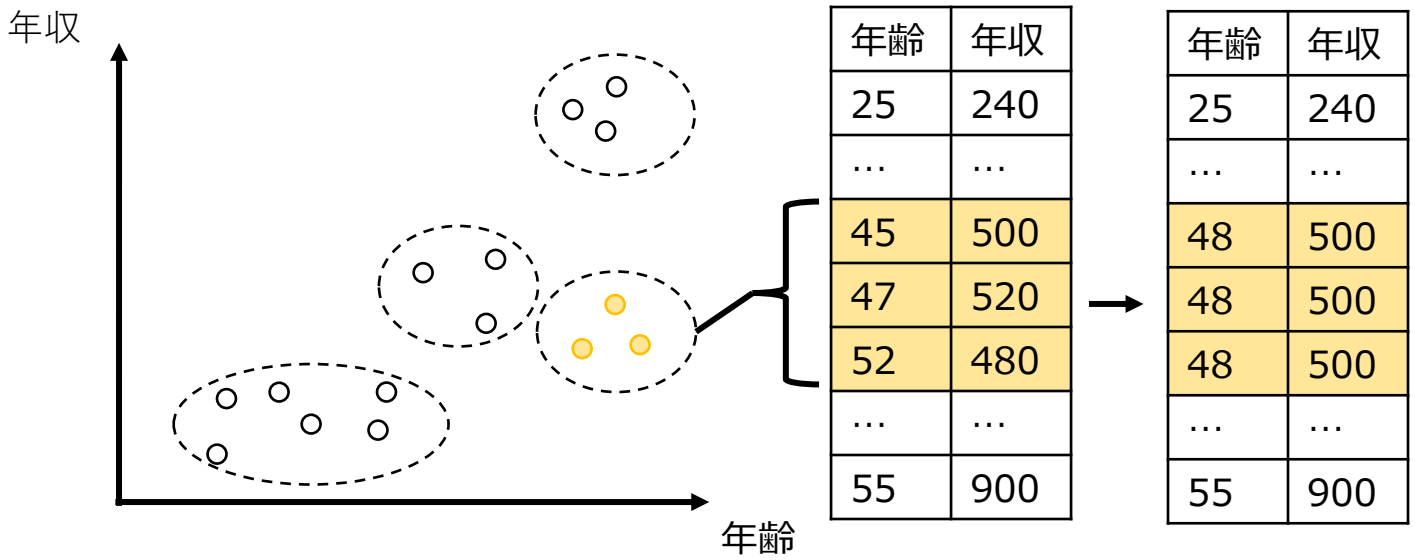
ー属性抑制：k-匿名性を満たさない原因となる属性を削除する

マイクロアグリゲーション

・ある数値属性についてレコードを複数のグループに分け、各グループのその数値属性の値をそのグループの代表値に置き換える

ー代表値としては平均値、中央値などが考えられる

ー通常のクラスタリング同様、グループを分けるルールは様々な方法が考えられる



年齢	年収
25	240
…	…
45	500
47	520
52	480
…	…
55	900

年齢	年収
25	240
…	…
48	500
48	500
48	500
…	…
55	900

データ匿名化手法：匿名化

k-匿名化データの有用性

・k-匿名性を持たすために行う加工処理によりデータの粒度が粗くなる

・k-匿名化されたデータの有用性を評価する指標が必要

・k-匿名化されたデータの有用性は分析目的が決まっていないこともあるため、データの変化量で表されることが多い

ー数値属性はユークリッド距離、カテゴリ属性はカルバックライブラーダイバージェンスなどで評価される

・D, D'：匿名化前後のデータセット

・x_{ij}, x'_{ij}：匿名化前後のi番目のレコードの属性jの属性値

・p_j, p'_j：匿名化前後の属性jの確率的質量分布

$$d_{\mathrm{Euclid}}(D, D') = \sum_{i=1}^{n} \sum_{j=1}^{d} \left(x_{ij} - x'_{ij}\right)^2$$

$$d_{KL}(D, D') = \sum_{j=1}^{d} KL\left(p_j, p'_j\right) = \sum_{j=1}^{d} \sum_{x_j \in X_j} p_j(x) \cdot \log \frac{p_j(x)}{p'_j(x)}$$

・その他、分析目的が決まっている場合は匿名化前後のデータを用いた際の分析結果の差などで評価する

データ匿名化手法：匿名化

履歴データの仮名化/匿名化

・ここまで一人の情報が一つのレコードに含まれる個人属性データを扱った

・履歴データにおける特定リスクを評価するには、ある個人を表す複数のレコードの集合に対するリスクを評価する必要がある

　ー購買履歴においては、ある個人のデータは{仮名ID, (日付情報、商品、数、価格), … (日付情報、商品、数、価格)}となる

・同一個人を長期にわたって観測できると、その個人の特定リスクは上昇する

・したがって、一定期間ごとに仮名IDを変更する必要がある

　ーk-匿名性の観点からも、同一個人に関する属性数が増えるに従い、k-匿名性の達成は難しくなる（**次元の呪い**）

・履歴データについては特異性、習慣性、一意性を考慮する必要があり、よりプライバシ保護が難しくなる

　ー現実的には**技術と法制度の両面からプライバシを守る必要がある**

k-匿名化まとめ

k-匿名化データ

・k-匿名性は実際に利用されているプライバシ指標であり、わかりやすく、匿名化も比較的容易

・一方で予期しない攻撃に対しては脆弱であり、それを補うような様々な指標が提案されている

・有用性は匿名化前後のデータを使って評価される

差分プライバシ導入

識別不可能性と攻撃者モデル

計算と秘匿性

・秘密にしたい入力xについてある関数fの出力$y = f(x)$を公開することを考える

・攻撃者はyと背景知識をもとにxを推測する

　ー背景知識はx以外のすべての情報

・このときxがどの程度推測されるかを評価したい

　ー暗号化アルゴリズム：暗号文$f(x)$から秘密鍵を使わずに平文xがどの程度推測されるか

　ー匿名化アルゴリズム：匿名化データ$f(x)$から元データxに関する情報がどの程度推測されるか

　ー統計分析：統計解析の結果$f(x)$から元データxに関する情報がどの程度推測されるか

　ー秘密計算：プロトコルの出力$f(x)$から、プロトコルの途中で得られる情報を背景知識として元データxに関する情報がどの程度推測されるか

識別不可能性と攻撃者モデル

識別不可能性

・関数fの秘匿性は識別不可能性と呼ばれる概念にもとづく

・識別不可能性とは異なる入力x, x'に対して$f(x)$と$f(x')$の見分けることの難しさを表す

－$f(x) = f(x')$となるとき、関数fは**完全秘匿性**を持つ

－$f(x), f(x')$を効率的な方法で見分けることが出来ないとき、関数fは**計算量的識別不可能性**を持つ

・効率的な方法で見分けがつかないとは、任意の$x, x' \in \{0,1\}^n$に対して以下が成り立つことをいう

$$\left|\Pr\left(A\left(f(x)\right) = 1\right) - \Pr\left(A\left(f(x')\right) = 1\right)\right| \leq \mathrm{negl}(n)$$

・ここで$A(\cdot)$は関数fの入力を正しく推測できたとき1を出力する攻撃アルゴリズム、$\mathrm{negl}(n)$はnに対して無視できる関数

識別不可能性と攻撃者モデル

<u>完全秘匿性をもつデータ匿名化</u>

・暗号化アルゴリズムに計算量的識別不可能性を持たせることは可能

ーデータ利用者は秘密鍵を持つため容易に復号可能

ー**秘密鍵を持たない攻撃者は**暗号化データ$f(x)$の**元の平文xと異なる平文x'との区別がつかない（xの情報を得られない）**

・匿名化や統計処理では完全秘匿性を持たせることは不可能

ーデータ利用者＝攻撃者

ー攻撃者が匿名化データ$f(x)$の元データxと異なるデータx'の違いが分からない（xの情報を得られない）
ということは、**データ利用者が匿名化データ$f(x)$からxの情報を全く得られない**ということ

・データベースに対するクエリを$q: \mathcal{D} \to Y$とすると、完全秘匿性を持つアルゴリズムfは以下のように定義できる

$$\forall D, D' \in \mathcal{D}, \forall S \subseteq Y, \Pr(f(q, D) \in S) = \Pr(f(q, D') \in S)$$

$$\Rightarrow \forall D, D' \in \mathcal{D}, \forall S \subseteq Y, \frac{\Pr(f(q, D) \in S)}{\Pr(f(q, D') \in S)} = 1$$

ー$f(q, D)$はDを入力としたクエリqの応答に対して、fを適用した値

・一方で秘匿性が全くないようなアルゴリズムfは以下で表される

$$\forall D, D' \in \mathcal{D}, \forall S \subseteq Y, f(q, D) \neq f(q, D')$$

$$\Rightarrow \forall D, D' \in \mathcal{D}, \forall S \subseteq Y, \frac{\Pr(f(q, D) \in S)}{\Pr(f(q, D') \in S)} = \infty$$

ークエリ出力$s \in S$から入力Dが確率1で推測できる場合がある（k-匿名化データは識別不可能性においては秘匿性は全くない）

差分プライバシ導入まとめ

識別不可能性に基づくプライバシ

・正当なユーザが攻撃者になりうるため、匿名化については完全秘匿性を持たせることは不可能

・識別不可能性に基づくと、k-匿名化データは秘匿性はないと見なされる

差分プライバシ

統計分析におけるプライバシリスク

一般にプライバシ情報を第三者から守りながらデータ活用をしたい場合、データの解析結果のみを提供した方がプライバシの観点から問題が少ない

統計量の公開プロセス

１．データ提供者は個人からデータを収集する

２．データ利用者はデータ収集者に対して統計解析を要求する

３．データ提供者はクエリに応じた統計解析を実施し、その結果をデータ利用者に提供する

統計量の公開におけるリスク

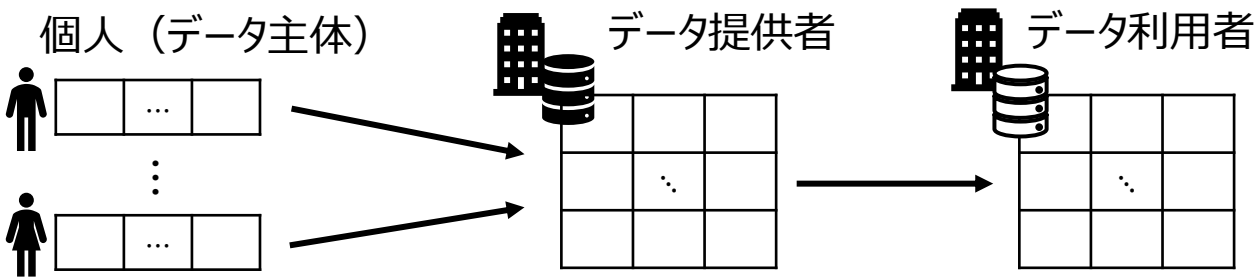
・統計量は多数の情報を用いて全体的な傾向を表すため、個人の情報の推測は困難に思える

・しかし条件によっては公開された統計量から個人に関する情報を推定することが可能

　ー極端な例では個人の情報が分かる

・複雑な設定であっても、統計量には個人の情報が少なからず含まれている

　ー個人の情報がどの程度漏れるかを定量的に評価する必要がある



年齢	テスト
19	40
20	80
21	78
22	82

全員の平均：70点
20代の平均：80点

10代の合計点：
$70 \times 4 - 80 \times 3 = 40$

統計分析におけるリスク評価指標：差分プライバシ

弱秘匿性の実現

・$f(x)$から有用な情報を残しつつxに関する情報を推測されないようにするために、弱い秘匿性を導入する

・データベースに対するクエリを$q: \mathcal{D} \to Y$とすると、弱秘匿性を持つアルゴリズムfは以下のように定義できる

定義：弱秘匿性

クエリ$q: \mathcal{D} \to Y$において、以下を満たす確率的アルゴリズムfは弱秘匿性を持つ。

$$\forall D, D' \in \mathcal{D}, \forall S \subseteq Y, \frac{\Pr(f(q, D) \in S)}{\Pr(f(q, D') \in S)} \leq c$$

ここで、cは1より大きい定数である。

・つまり弱秘匿性を実現するためには真のクエリ応答値に確率的な揺らぎを与えればよい

ー以後この確率的な揺らぎを与える関数を確率的メカニズムMとする

定数cの与え方

・有用性を維持するためには確率的メカニズムの出力は真のクエリ応答値に近い必要がある

ーD, D'が似ているほど確率的メカニズムの出力は近くなる必要がある

ーデータベースの距離、例えばDとD'において同一でないレコードの数、を$d(D, D')$とする

・秘匿性の強弱をパラメータϵで与えると、例えば$c = \exp\left(\epsilon \cdot d(D, D')\right)$とすることができる

統計分析におけるリスク評価指標：差分プライバシ

定義：ϵ-差分プライバシ

クエリ$q \in Q$において、$d(D, D') = 1$であるような任意のデータベース$D, D' \in \mathcal{D}$、および任意の出力の部分集合$S \subseteq Y$について

$$\Pr(M(q, D) \in S) \leq \exp(\epsilon) \cdot \Pr(M(q, D') \in S)$$

$$\Leftrightarrow \frac{\Pr\big(M(q, D)\big)}{\Pr\big(M(q, D')\big)} \leq \exp(\epsilon)$$

であるとき、確率的メカニズムMはϵ-差分プライバシを満たす。ただし$\epsilon \geq 0$とする。

定理：差分プライバシが保証する秘匿性

差分プライバシは弱秘匿性を保証する

証明

・$d(D_0, D_m) = m$となるD_0, D_mを想定する

・このとき$d(D_0, D_1) = d(D_1, D_2) = \cdots = d(D_{m-1}, D_m) = 1$となるデータベース$D_1, D_2, \cdots D_{m-1}$が存在する

・Mがϵ-差分プライバシを満たす時、任意の$i \in \{0, \cdots, m\}$に対して以下が成り立つ

$$\frac{\Pr\big(M(q, D_i)\big)}{\Pr\big(M(q, D_{i+1})\big)} \leq \exp(\epsilon)$$

・したがって

$$\frac{\Pr\big(M(q, D_0)\big)}{\Pr\big(M(q, D_1)\big)} \cdot \frac{\Pr\big(M(q, D_1)\big)}{\Pr\big(M(q, D_2)\big)} \cdot \cdots \cdot \frac{\Pr\big(M(q, D_{m-1})\big)}{\Pr\big(M(q, D_m)\big)} = \frac{\Pr\big(M(q, D_0)\big)}{\Pr\big(M(q, D_m)\big)} \leq \exp(\epsilon \cdot m) = \exp\big(\epsilon \cdot d(D_0, D_m)\big) = c$$

統計分析におけるリスク評価指標：差分プライバシ

ϵの持つ意味

・ϵはプライバシの強さを表すパラメータである

ー隣接した任意のデータベースにおいて、**同じクエリの出力値を取る確率が高々$e^{\epsilon} \approx 1 + \epsilon$倍程度しか変わらない**ことを意味する

ーϵが小さいほど強いプライバシを保証する

・差分プライバシの定義ではデータベースの隣接性（$d(D, D') = 1$）は明確に定義されておらず、個別にプライバシの保護対象となる情報に応じて定義する必要がある

（e.g.）SNSにおけるユーザ同士の関係のプライバシ

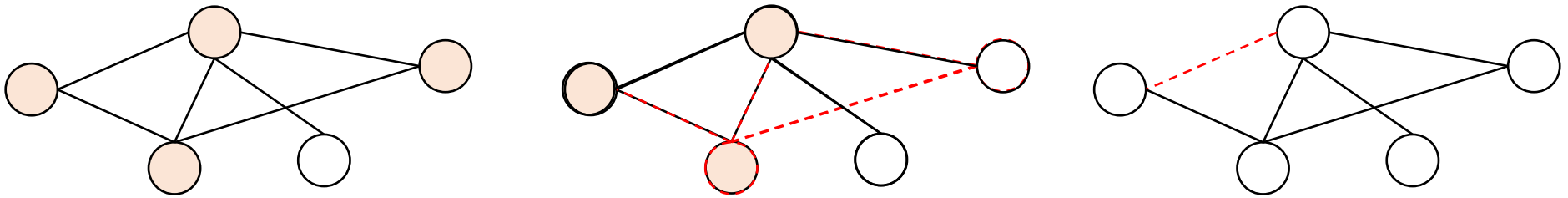
・各頂点をユーザ、各エッジをフォローの有無を表すようなグラフ$G = (V, E)$を想定し、このグラフ全体をデータベースとして捉える

・グラフ$G = (V, E)$における隣接性はユーザ、あるいはエッジに着目して考えられる

・クエリを「Gにおいてフォロー数2人以上のユーザ数」とすると、それぞれの隣接性が持つプライバシの意味は以下のようになる

ーユーザ：あるユーザがSNSに参加しているかどうかにかかわらず、「Gにおいてフォロー数2人以上のユーザ数がkである」と出力される確率はe^{ϵ}倍程度しか差がない

ーエッジ：あるユーザがSNSの他の1人をフォローしているかに関わらず、「Gにおいてフォロー数2人以上のユーザ数がkである」と出力される確率はe^{ϵ}倍程度しか差がない



統計分析におけるリスク評価指標：差分プライバシ

定義：(ϵ, δ)-差分プライバシ

クエリ$q \in Q$において、$d(D, D') = 1$であるような任意のデータベース$D, D' \in \mathcal{D}$、および任意の出力の部分集合$S \subseteq Y$について、

$$\Pr(M(q, D) \in S) \leq \exp(\epsilon) \cdot \Pr(M(q, D') \in S) + \delta$$

であるとき、確率的メカニズムMは(ϵ, δ)-差分プライバシを満たす。ただし$\epsilon, \delta \geq 0$とする。

δの持つ意味

・(ϵ, δ)-差分プライバシはϵ-差分プライバシの緩和版であるといえる

－$\delta = 0$のときϵ-差分プライバシと等価であり、δが大きくなるほどMは確率的にϵ-差分プライバシを満たさなくなる

・(ϵ, δ)-差分プライバシを満たすメカニズムは$1 - \frac{2\delta}{e^{\epsilon}\epsilon}$の確率で$2\epsilon$-差分プライバシを保証する

－$\frac{2\delta}{e^{\epsilon}\epsilon}$の確率で珍しい値を持つレコード（外れ値）に関する情報が漏れていると考えられる

定理：(ϵ, δ)-差分プライバシを満たすメカニズムのϵ-差分プライバシにおけるプライバシ保証

(ϵ, δ)-差分プライバシを満たすメカニズムMについて、少なくとも確率$1 - \delta'$で以下が成り立つ

$$\Pr(M(D) = y) \leq e^{\epsilon'} \Pr(M(D') = y)$$

ここで$\epsilon' = 2\epsilon, \delta' = \frac{2\delta}{e^{\epsilon}\epsilon}$である。

統計分析におけるリスク評価指標：差分プライバシ

定理：(ϵ, δ)-差分プライバシを満たすメカニズムのϵ-差分プライバシにおけるプライバシ保証

(ϵ, δ)-差分プライバシを満たすメカニズムMについて、少なくとも確率$1-\delta'$で以下が成り立つ

$$\Pr(M(D)=y) \leq e^{\epsilon'} \Pr(M(D')=y)$$

ここで$\epsilon' = 2\epsilon, \delta' = \frac{2\delta}{e^{\epsilon}\epsilon}$である。

証明

・2ϵ-差分プライバシを満たさないメカニズムの出力値の集合を$Z = \{y | \Pr(M(D)=y) \geq e^{2\epsilon} \Pr(M(D')=y)\}$とすると

$$\Pr(M(D) \in Z) \geq e^{2\epsilon} \Pr(M(D') \in Z) \geq e^{\epsilon}(1+\epsilon) \Pr(M(D') \in Z) \ (\because e^{\epsilon} \geq 1+\epsilon)$$

$$\Leftrightarrow \Pr(M(D) \in Z) - e^{\epsilon} \Pr(M(D') \in Z) \geq e^{\epsilon}\epsilon \Pr(M(D') \in Z)$$

$$\Leftrightarrow e^{\epsilon}\epsilon \Pr(M(D') \in Z) \leq \Pr(M(D) \in Z) - e^{\epsilon} \Pr(M(D') \in Z) \leq \delta \ (\because M: (\epsilon, \delta) - \text{差分プライバシ})$$

$$\Rightarrow \Pr(M(D') \in Z) < \frac{\delta}{\epsilon e^{\epsilon}}$$

・以上より$1 - \frac{\delta}{\epsilon e^{\epsilon}}$の確率で$\Pr(M(D)=y) \leq e^{2\epsilon} \Pr(M(D')=y)$、すなわち$2\epsilon$-差分プライバシが成り立つ

統計分析におけるリスク評価指標：差分プライバシ

差分プライバシにおける攻撃者の背景知識

・k-匿名性において、攻撃者はDに含まれる間接識別情報の組み合わせを背景知識として持つ

　－l-多様性やその他の指標ではその他要配慮属性の組み合わせ

・差分プライバシにおいては、攻撃者は任意の背景知識を持つ

　－入力データベースを n bit列とした時、攻撃者は$n-1$ bitを知っており、残りの1bitについて推測するケースも含む

　－差分プライバシはプライバシの下限を保証するものであり、非常に強力な安全性指標といえる

統計分析におけるリスク評価指標：差分プライバシ

差分プライバシにおけるメカニズムMのプライバシと有用性

・プライバシ：ϵをパラメータとしてϵ-差分プライバシを保証する

・有用性：差分プライバシメカニズムの有無の差分

ー実験的には差分プライバシメカニズムを適用することでクエリ応答値がどの程度真の値から離れているかを見る

ー理論的にはメカニズムの適用の有無による統計量の差が発生する確率の上界を考える

$$\Pr\big(\|q(D) - M(q, D)\| > g(n)\big) < \beta$$

・ここで$g(n)$はレコード数nに依存した出力の収束の性質

・例えば$g(n) = 1000/n \in O(1/n)$であればレコード数を10倍すれば、メカニズムと真の出力差を1/10にできる

ー差分プライバシではレコード数nが性能に大きな影響を与える

・差分プライバシメカニズムによるノイズの影響が$g(n)$を超える確率はβより小さいことを保証

⇔確率$1-\beta$で差分プライバシメカニズムによるノイズは$g(n)$よりも小さいことを保証

差分プライバシまとめ

差分プライバシ

・識別不可能性に基づき、弱い秘匿性を持つプライバシ指標

・差分プライバシにおいては攻撃者の背景知識は問わない

・確率的な揺らぎを与えることで差分プライバシが満たされる

差分プライバシメカニズム

差分プライバシを満たす確率的メカニズム

敏感度（sensitivity）の導入

・数値属性$x_i \in \mathbb{R}$からなるデータベース$D = \{x_1, \cdots, x_n\}$とクエリ$q: \mathbb{R}^n \to \mathbb{R}$を考える

・$d(D, D') = 1$であるようなデータベースの組$D = \{x_1, \cdots, x_n\}, D' = \{x_1, \cdots, x_{n-1}, x'_n\}$を隣接データベースとよぶ

・1レコード異なるデータベースにクエリ出力が与える影響の大きさを敏感度（sensitivity）とよぶ

ー敏感度は以下で定義される。ただし$\|\cdot\|_p$はl_pノルムである。

$$\Delta_{p,q} = \max_{\forall D, D' : d(D,D')=1} \|q(D) - q(D')\|_p$$

敏感度の具体例

・敏感度はクエリに応じて大幅に異なる

・レコードの定義域を$x \in [0, m]$とする

ー頻度関数 $q_{1,frequency} = \max_{k \in [0,m]} |k - (k-1)| = 1$（$k$は$x_n$の属性値の頻度）

ー平均関数 $q_{1,average} = \frac{1}{n} \max_{x_i, x'_i \in [0,m]} |x_n - x'_n| = \frac{m}{n}$

ー最大値関数 $q_{1,max} = \max_{x_i, x'_i \in [0,m]} |x_n - x'_n| = m$

差分プライバシを満たす確率的メカニズム

ラプラスメカニズム

・ラプラスメカニズム：真のクエリの出力にラプラス分布から生成した乱数を加えるメカニズム

ーラプラス分布：$Lap(R) = \frac{1}{2R} e^{-\frac{|x|}{R}}$

ラプラスメカニズムのアルゴリズム

1．データベースD、プライバシパラメータϵ、クエリqの敏感度$\Delta_{1,q}$を入力する

2．ラプラス分布$Lap\left(\frac{\Delta_{1,q}}{\epsilon}\right)$に従う確率変数$r$を計算する

3．$y = q(D) + r$を出力する

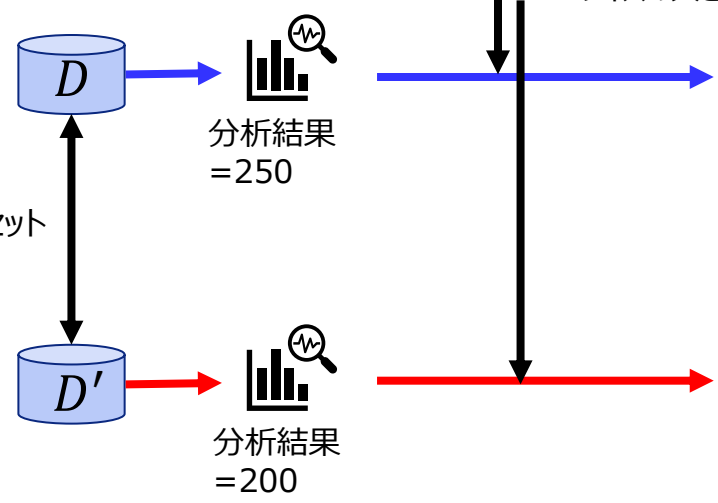
ラプラスメカニズムのイメージ

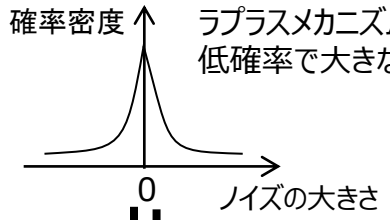
クエリ：40歳以上の貯金の平均値

名前	年齢	貯金
青木	33	1,000
⋮	⋮	⋮
山田	47	200
湯川	49	300

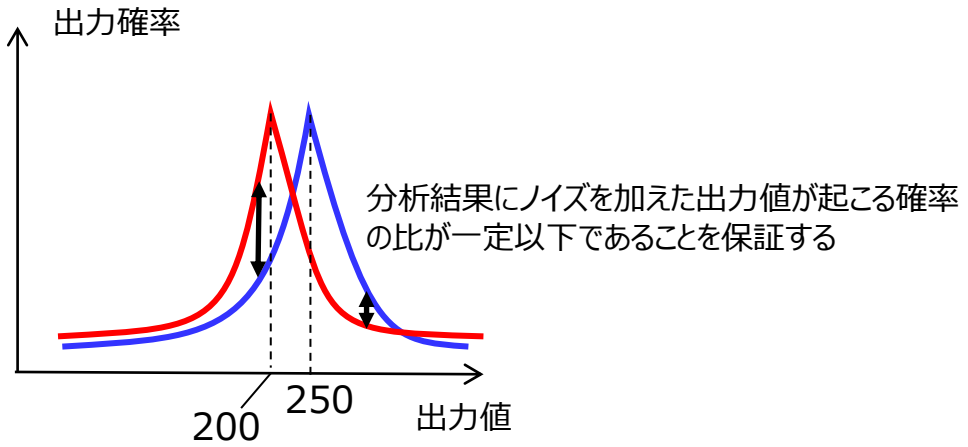
1レコード異なるデータセット

名前	年齢	貯金
青木	33	1,000
⋮	⋮	⋮
山田	47	200





ラプラスメカニズム：高確率で小さな値（0付近）、
低確率で大きな値となるノイズを分析結果に加える



分析結果にノイズを加えた出力値が起こる確率
の比が一定以下であることを保証する

差分プライバシを満たす確率的メカニズム

定理：ラプラスメカニズムのプライバシ保証

ラプラスメカニズムはϵ-差分プライバシを満たす。

証明

・ラプラスメカニズムの応答値の確率密度分布は以下のように与えられる

$$p\big(M(D,q)=y(=q(D)+r)\big)=\frac{1}{2R}\exp\left(-\frac{|r|}{R}\right)=\frac{\epsilon}{2\Delta_{1,q}}\cdot\exp\left(-\frac{\epsilon|y-q(D)|}{\Delta_{1,q}}\right)$$

・したがってラプラス分布における確率密度の比は以下の通り

$$\begin{aligned}\left|\frac{p(M(D,q)=y)}{p(M(D',q)=y)}\right| &= \left|\frac{\exp\left(-\frac{\epsilon|y-q(D)|}{\Delta_{1,q}}\right)}{\exp\left(-\frac{\epsilon|y-q(D')|}{\Delta_{1,q}}\right)}\right| \\ &= \left|\exp\left(\epsilon\cdot\frac{|y-q(D)|-|y-q(D')|}{\Delta_{1,q}}\right)\right| \\ &\le \exp\left(\epsilon\cdot\frac{|q(D)-q(D')|}{\Delta_{1,q}}\right)(\because \text{三角不等式}) \\ &\le \exp\left(\epsilon\cdot\frac{\Delta_{1,q}}{\Delta_{1,q}}\right)=e^{\epsilon}\end{aligned}$$

・ラプラスメカニズムはyを出力する確率の比を常にe^{ϵ}以下に抑えられるため、ϵ-差分プライバシを満たす

差分プライバシを満たす確率的メカニズム

定理：ラプラスメカニズムの理論的有用性

ラプラスメカニズムは任意の$\beta \in (0,1]$について、以下が成り立つ。

$$\Pr\left(\|y - q(D)\|_1 > \frac{\Delta_{1,q}}{\epsilon} \cdot \ln\frac{1}{\beta}\right) \leq \beta$$

証明

・ラプラス分布に従う確率変数をrとすると、ラプラス分布の裾の確率は以下のように与えられる

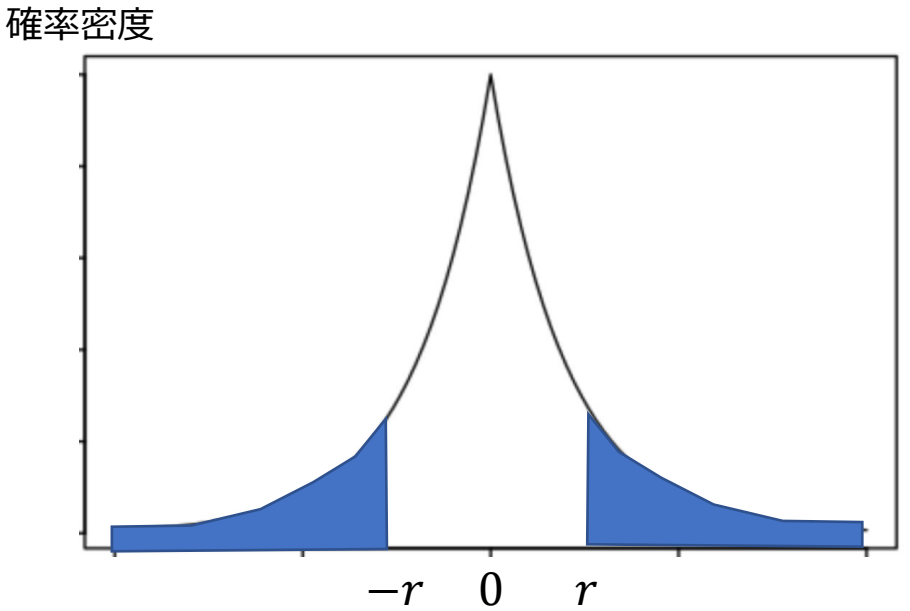
$$\Pr(|r| > t \cdot R) = \exp(-t)$$

$$\Leftrightarrow \Pr\left(\|y - q(D)\|_1 > \frac{\Delta_{1,q}}{\epsilon} \cdot \ln\frac{1}{\beta}\right) = \beta$$

プライバシと有用性のトレードオフ

・定理より、$\frac{\Delta_{1,q}}{\epsilon}$が大きいほど、誤差が大きくなる確率が高まることが分かる

・プライバシの強さと有用性はトレードオフの関係にある



差分プライバシを満たす確率的メカニズム

ラプラスメカニズムの具体例：年齢

・年齢の定義域を[0,100]、データベースのサイズを$n = 100$、プライバシパラメータを$\epsilon = 0.1$とする

ークエリが平均値の場合、$\Delta_{1,ave} = \frac{100-0}{n} = \frac{100}{n} = O\left(\frac{1}{n}\right)$

ークエリが最大値の場合、$\Delta_{1,max} = 100 - 0 = 100 = O(1)$

$$\Pr\left(\|y - q(D)\|_1 > \frac{\Delta_{1,q}}{\epsilon} \cdot \ln\frac{1}{\beta}\right) \leq \beta$$

・平均値を出力とするラプラスメカニズムは95%の確率で真の値との誤差が$\frac{100}{\epsilon n}\ln\frac{1}{\beta} = \frac{1}{0.1}\ln\frac{1}{0.05} \approx 3$以下であることが保証される

ー$n = 100{,}000$まで増加させると、95%の確率で$\frac{100}{\epsilon n}\ln\frac{1}{\beta} = \frac{1}{0.1\times 100}\ln\frac{1}{0.05} \approx 0.03$以下であることが保証される

・同様に最大値を出力とするラプラスメカニズムは95%の確率で$\frac{100}{\epsilon}\ln\frac{1}{\beta} = \frac{100}{0.1}\ln\frac{1}{0.05} \approx 2995.7$以下であることが保証される

ー$O(1)$なので、nをいくら増やしても精度の向上は見込めない

ー80%の確率でもノイズは1609.4以下であることしか保証できない

クエリと敏感度

・クエリによっては統計解析として問題のある結果となる

・差分プライバシを満たすメカニズムはラプラスメカニズムの他、指数メカニズム、ガウシアンメカニズムなどが提案されている

・差分プライバシを満たすために、敏感度を下げるための様々な研究がある

ー具体的な手法は次回講義

差分プライバシメカニズムまとめ

差分プライバシメカニズム

・差分プライバシを満たすための確率的アルゴリズムとして、ラプラスメカニズムや指数メカニズムが存在する

・各メカニズムを設計するにあたり、クエリに応じた敏感度の定義が必要

・クエリによっては敏感度が大きく、非常に大きなノイズが付与されうる

局所差分プライバシ

匿名化モデル・差分プライバシモデルの前提

匿名化モデルや差分プライバシモデルでは各個人から正確なデータを収集した信頼された機関が、第三者機関にデータを提供する

データ提供のプロセス

1．データ提供者（信頼された機関）は個人からデータを収集する

2．データ提供者（信頼された機関）は、収集したデータを加工し、データ利用者（第三者機関）に提供する

3．データ利用者（第三者機関）は、提供されたデータを用いてデータ解析を行う

統計量の公開プロセス

1．データ提供者（信頼された機関）は個人からデータを収集する

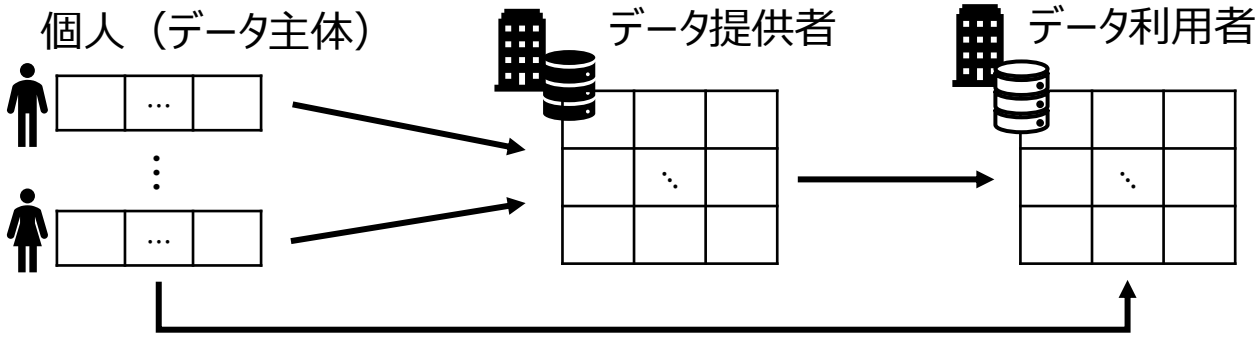
2．データ利用者（第三者機関）はデータ収集者に対して統計解析を要求する

3．データ提供者（信頼された機関）はクエリに応じた統計解析を実施し、その結果をデータ利用者（第三者機関）に提供する

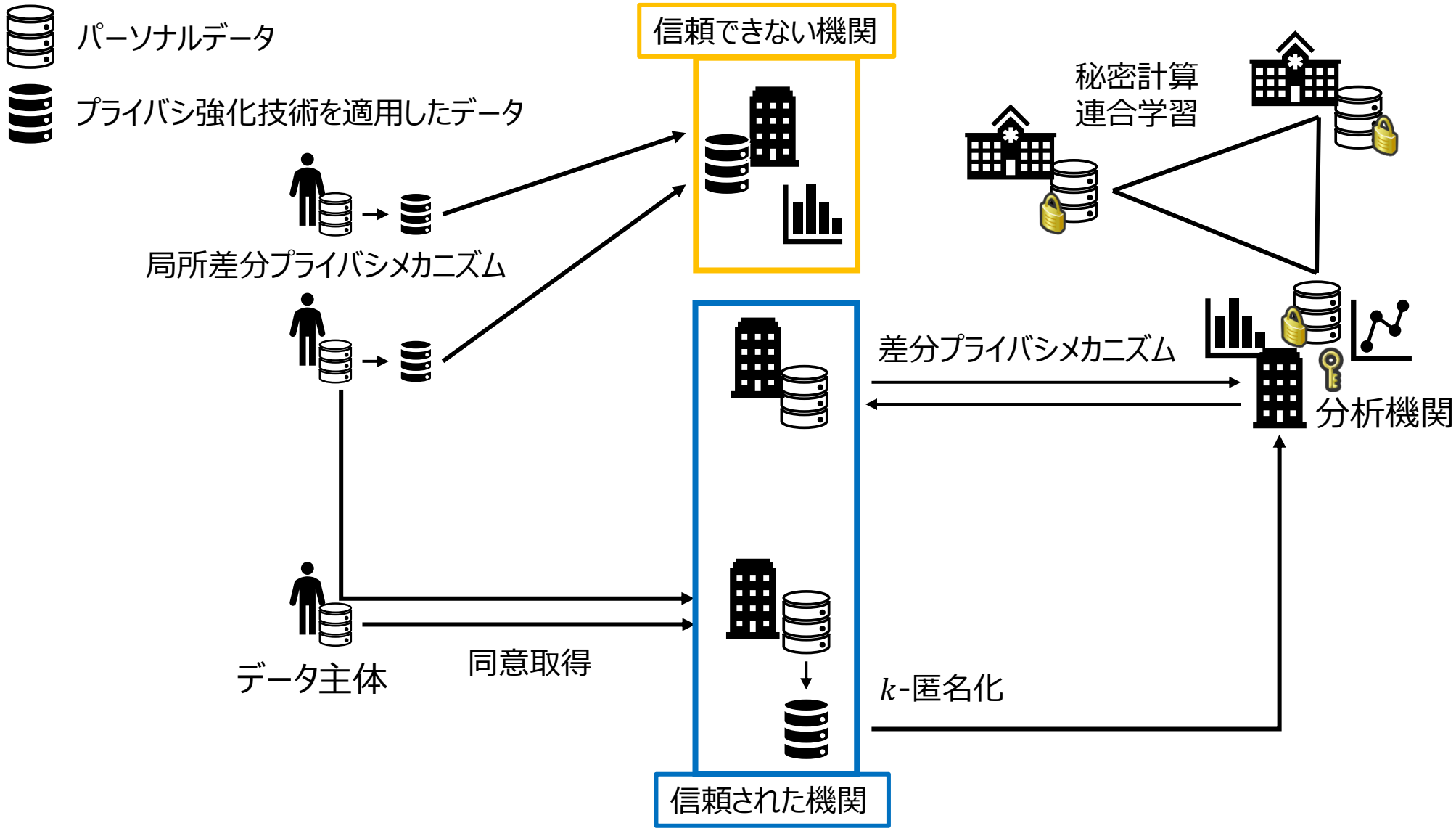
信頼できない機関にデータ提供する際のプロセス

1．個人がデータを加工する

2．個人がデータ利用者にデータを提供する



匿名化モデル・差分プライバシモデルの前提

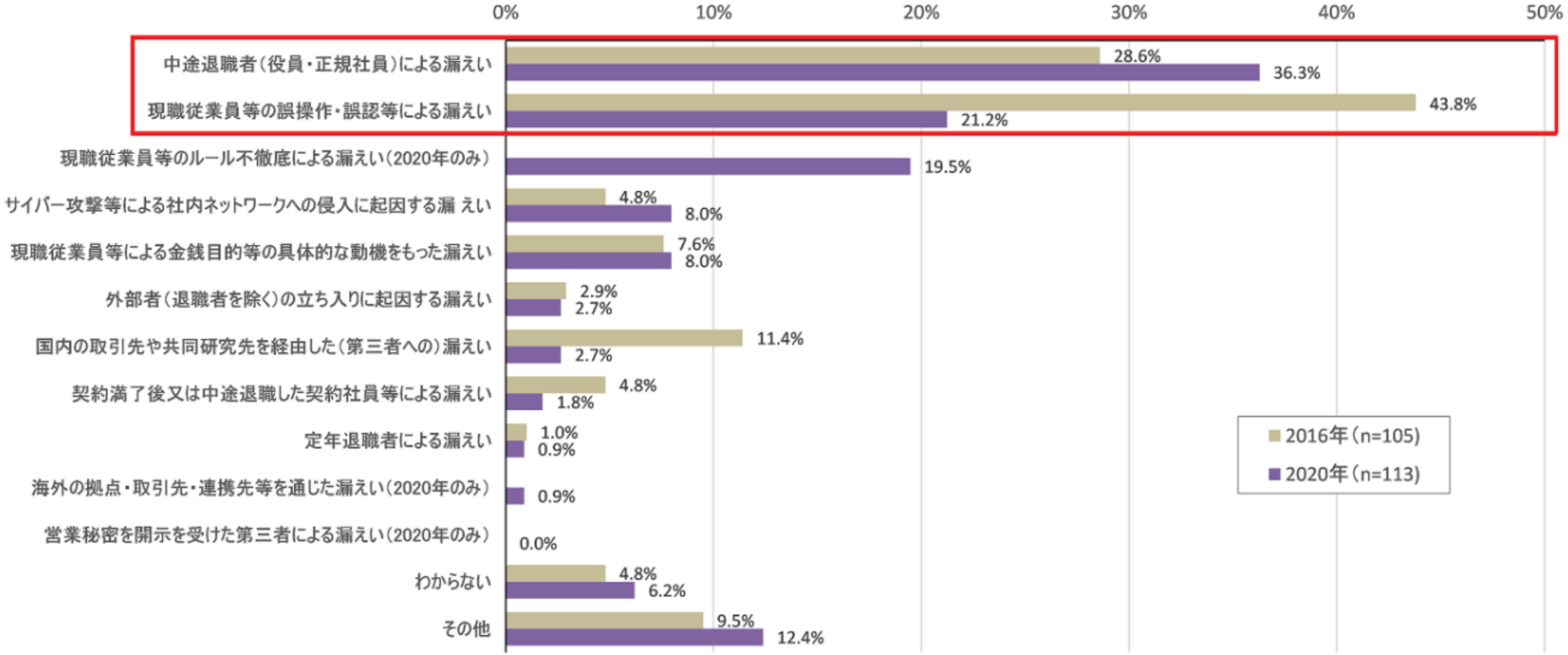


匿名化モデル・差分プライバシモデルの前提

<u>「信頼できるはず」の機関も信頼できない可能性がある</u>

・2020年における情報漏えいに関するインシデントの原因のトップは「中途退職者」による漏えい

・故意でなくとも情報が漏洩することもあり、パーソナルデータを企業に預けること自体に抵抗感を感じる人も



IPA「企業における営業秘密管理に関する実施調査2020」
https://www.ipa.go.jp/security/fy2020/reports/ts_kanri/index.html

個人のデータ提供におけるリスク評価指標：局所差分プライバシ

定義：ϵ-局所差分プライバシ

任意のデータ$x, x' \in X$において、任意の出力の部分集合$S \subseteq Y$について

$$\Pr(M(x) \in S) \leq \exp(\epsilon) \cdot \Pr(M(x') \in S)$$

$$\Leftrightarrow \frac{\Pr\left(M(x)\right)}{\Pr\left(M(x')\right)} \leq \exp(\epsilon)$$

であるとき、確率的メカニズムMはϵ-局所差分プライバシを満たす。ただし$\epsilon \geq 0$とする。

ϵの持つ意味

・ϵはプライバシの強さを表すパラメータである

ー任意のデータにおいて、出力データが一致する確率が高々$e^{\epsilon} \approx 1 + \epsilon$倍程度しか変わらない

ーϵが小さいほど強いプライバシを保証する

・局所差分プライバシは差分プライバシの特殊系として考えることができる

ー個人＝データ提供者と考えると、個人のデータ（データセット）の統計情報（例えばデータそのもの）を提供すると見なせる

ー差分プライバシにおける敏感度をデータの定義域とすれば、差分プライバシメカニズムは局所差分プライバシを保証する

（e.g.）$x \in [0,100]$において、敏感度を$\Delta_{1,q} = 100$、クエリを$q(x) = x$を出力するラプラスメカニズムはϵ-局所差分プライバシを保証する

個人のデータ提供におけるリスク評価指標：局所差分プライバシ

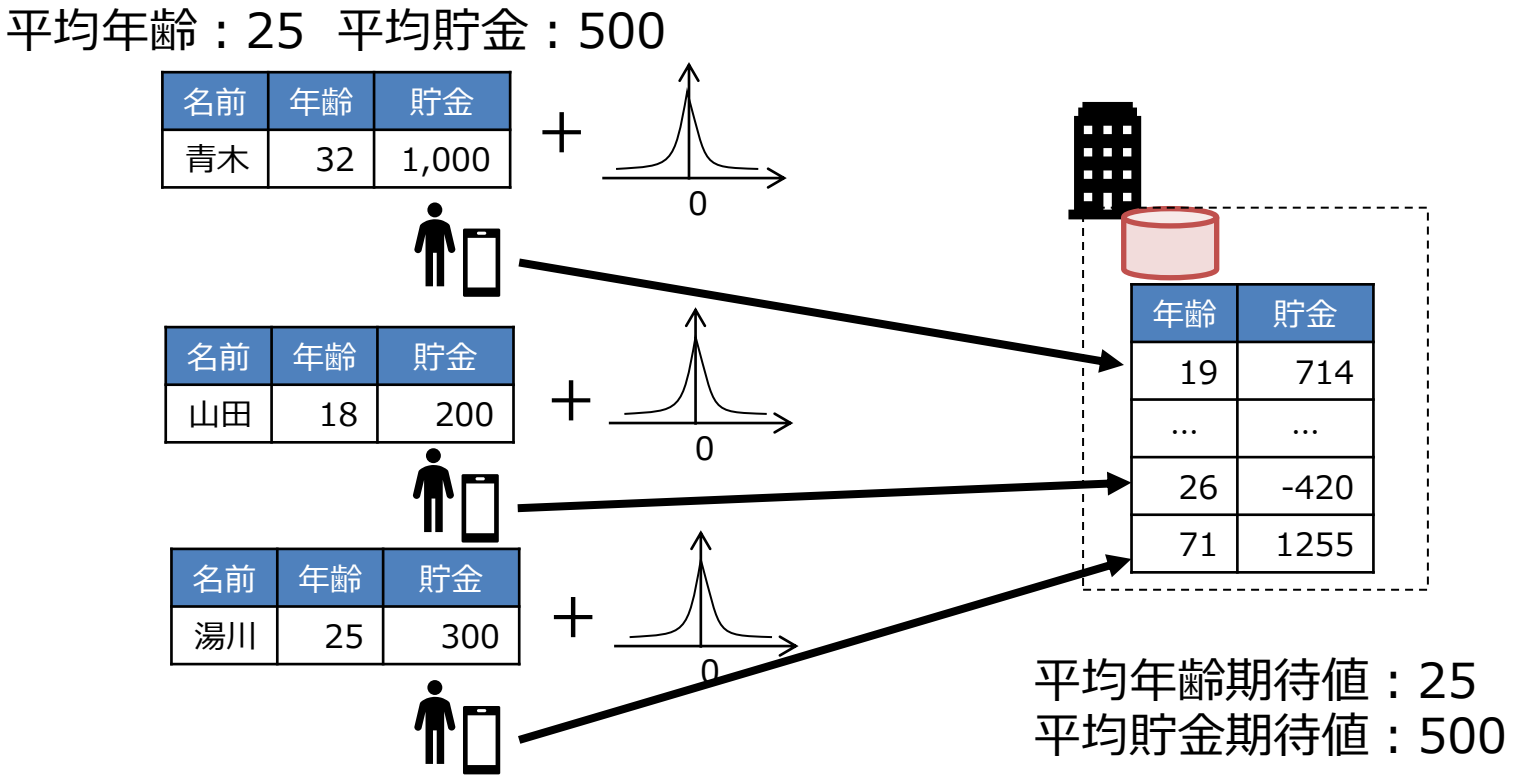
ラプラスメカニズムを用いた局所差分プライバシメカニズムのイメージ

・年齢であれば定義域は[0,100]程度なので、$\Delta_{1,q} = 100$

・貯金の場合、非常に広い定義域を考える必要があり、同様に敏感度が大きくなる

・理論的には局所差分プライバシを保証するが、出力値は真の値から大幅に外れることとなる
　ー場合によっては出力値が定義域を超える値にもなりうる

・各レコードには大きなノイズが付与される一方で、付与されるノイズの大きさの平均値は0
　ー局所差分プライバシメカニズムを適用したレコードの平均値は、レコード数が増えるにしたがって元のレコード全体の平均値に近づく
　ー多くの局所差分プライバシメカニズムでは各個人のデータを集めるが、特定のユースケースに着目した作りとなっており、個々のデータにはあまり意味がない



局所差分プライバシを満たすメカニズム

ランダム化応答

・ランダム化応答：二値の値を確率的に変えて出力するメカニズム

ランダム化応答のアルゴリズム

1．データ$x \in \{0,1\}$、プライバシパラメータϵを入力する

2．一様ランダムに$r \in [0,1]$を決定する

3．$r \leq \frac{\exp(\epsilon)}{1+\exp(\epsilon)}$であれば$y = x$を出力し、それ以外は$y = 1 - x$を出力する

定理：ランダム化応答のプライバシ保証

ランダム化応答はϵ-局所差分プライバシを満たす。

局所差分プライバシを満たすメカニズム

定理：ランダム化応答のプライバシ保証

ランダム化応答はϵ-局所差分プライバシを満たす。

証明

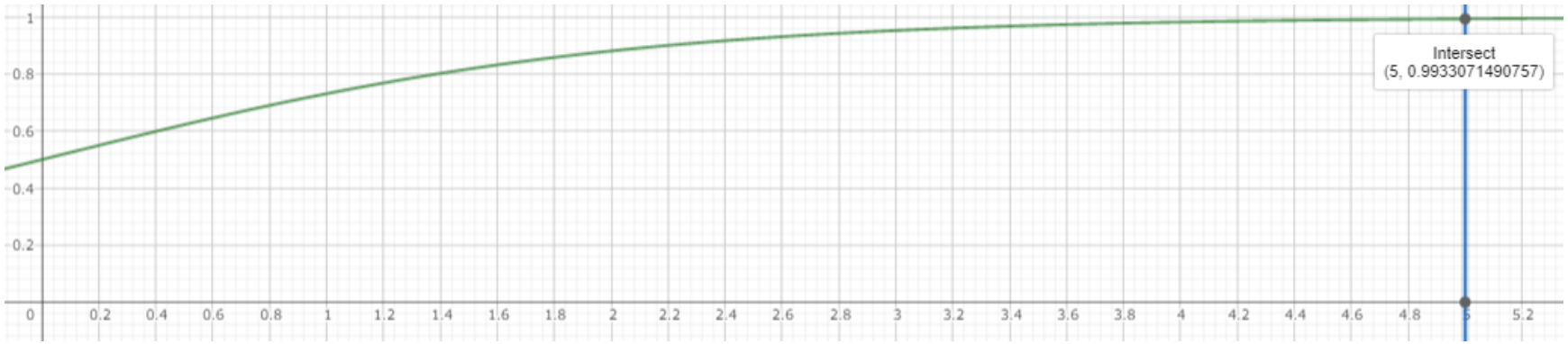
$\bar{x} = 1 - x$とすると、$\Pr(y = x|x) = \frac{e^{\epsilon}}{1+e^{\epsilon}}, \Pr(y = \bar{x}|x) = \frac{1}{1+e^{\epsilon}}$となる。したがって任意の$x, x' \in \{0,1\}$に対して

$$\frac{\Pr(y|x)}{\Pr(y|x')} \leq \frac{\Pr(x|x)}{\Pr(\bar{x}|x)} = \frac{\frac{e^{\epsilon}}{1+e^{\epsilon}}}{\frac{1}{1+e^{\epsilon}}} = \exp(\epsilon)$$

ϵの大きさによる直感的なランダム化の影響

・ランダム化応答は「あなたは過去に罪を犯したことがありますか？」などの機微な問に対して、正しく回答する（Yes or No）か嘘をつく（No or Yes）かというメカニズム

ϵ	$\Pr(x\|x)$
0	0.5
0.1	0.525
0.5	0.622
0.75	0.679
1.0	0.731
1.5	0.818
5	0.993



正しく回答する確率

局所差分プライバシを満たすメカニズム

一般ランダム化応答

・一般ランダム化応答：三値以上の値を確率的に変えて出力するメカニズム

一般ランダム化応答のアルゴリズム

1．データ$x \in \{1, \cdots, K\}$、プライバシパラメータϵを入力する

2．一様ランダムに$r \in [0,1]$を決定する

3．$r \leq \frac{\exp(\epsilon)}{K-1+\exp(\epsilon)}$であれば$y = x$を出力し、それ以外は$y \neq x$をランダムに出力する

定理：一般ランダム化応答のプライバシ保証

一般ランダム化応答はϵ-局所差分プライバシを満たす。

証明

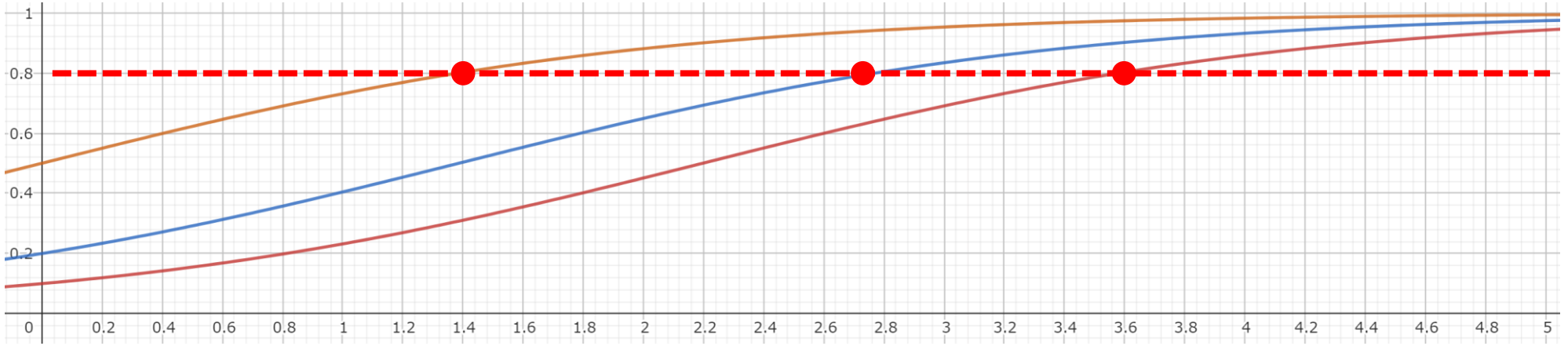
略

ϵの大きさによる直感的なランダム化の影響

・橙：$K = 2$

・青：$K = 5$

・赤：$K = 10$



局所差分プライバシまとめ

局所差分プライバシ

・差分プライバシと非常によく似たプライバシ指標として局所差分プライバシがある

・局所差分プライバシではユーザはデータ提供先を信頼する必要がなく、データ提供先も直接プライバシ情報を保有する必要がない

演習

演習課題

設定

・A銀行は保有する顧客情報を匿名加工情報に加工し、B社に販売したいと考えています

ーA銀行は顧客の定期預金のキャンペーン施策のために作成したデータをB社に提供予定です

ーB社は手広く様々なターゲティング広告を行う企業で、適切な広告配信のためにパーソナルデータを必要としています

・匿名化のノウハウがないA銀行はプライバシの専門家（講義参加者）に匿名加工メカニズム開発の依頼を検討しています

・A銀行は皆さんに顧客情報のサンプルデータを提供し、1ヶ月後（6/17）にコンペを開催予定です

ー想定されるリスクや有用性維持の方法、実際の匿名化メカニズムなどの提案を受けて、どのチームに開発を依頼するかを決定する予定です

演習課題

<u>演習の流れ</u>

・本日～5/31：匿名化フェーズ

－各チームXにそれぞれ異なる100レコードのデータセット（teamX_original.csv）を配布

－各チームはデータセットに対して匿名化処理を行って提出

・提出するデータは、**匿名化データセット（teamX_anonymized.csv）**

・6/1～6/11：攻撃フェーズ

－各チームXにそれぞれ異なる200レコード（うち100レコードはteamY_original.csvのレコード）のデータセット（teamY_candidate.csv）を配布

－各チームはデータセットに対してレコードの再識別攻撃を行って提出

・提出するデータは、**対応表（teamX_attack_teamY.csv）**

・6/17：発表

－各チームは15分間の発表（発表10分、質疑5分）を行う

・匿名化の方針（実装できなくてもよい）：想定リスク、有用性担保 etc.

・攻撃の方針（実装できなくてもよい）：どういった情報が取れるか、他チームの弱点 etc.

・実際の匿名化、攻撃のアルゴリズムの説明

・アピールポイントなど

演習課題

サンプルコード

ファイル	関数	入力	出力	説明
Anonymize_sample.py	Generalization(data, attr, l)	data(pandas.DataFrame形式)：データセット attr(list[string]形式)：一般化を行う属性のリスト l(int形式)：階層番号	一般化後のデータセット (pandas.DataFrame形式)	データをattr_depth.jsonに従って一般化する
	Remove_record(data, attr, k)	data(pandas.DataFrame形式)：データセット attr(list[string]形式)：匿名性を評価する属性 k(int形式)：匿名性を表すパラメータ	k-匿名性を保証したデータセット (pandas.DataFrame形式)	選択した属性において同じ属性値の組み合わせがk件以上のレコードを抽出する
	Differential_l(x, value_range, epsilon)	x(float形式)：数値データ単体 value_range(list[float]形式)：xが属する属性の全データ、または[最大値、最小値]のリスト epsilon(float形式)：ノイズの強さを表すパラメータ	ノイズ付加後のデータ (float形式)	局所差分プライバシを満たすラプラスメカニズム
	Differential_e(x, variation, epsilon)	x(float形式)：カテゴリデータ単体 value_range(list[string]形式)：xが属する属性の全データ、または[最大値、最小値]のリスト epsilon(float形式)：ノイズの強さを表すパラメータ	変換後のデータ (string形式)	局所差分プライバシを満たす一般ランダム化応答
exe_anonymize_code.py			匿名化データセット	匿名化のサンプル実行コード

演習課題

サンプルコード

・一般化はattr_depth.jsonに従う

カテゴリ属性の場合

```
“元の属性名”{
    “type” :”category”,
    “definition” : {
        “パラメータ名1” : {
            “depth_1” : “1階層のデータ名”,
            “depth_2” : “2階層のデータ名”
        },
        “パラメータ名2” : {
            ・・・
        }, ・・・
    }
}, ・・・
```

数値属性の場合

```
“元の属性名”{
    “type” :”numeric”,
    “threshold_1” : “閾値”,
    “threshold_2” : “閾値1/閾値2/閾値3”
}, ・・・
```

generarization()でl=1のとき、
カテゴリ属性は”depth_1”の値に置換する。
数値属性は”threshold_1”を境界に
データをグループ分け、
各グループの値をそのグループの中央値に置換する。

演習課題

サンプルコード

ファイル	関数	入力	出力	説明
Attack.py	Neighbor_att(ano_data, ata_data, att, n)	ano_data(pandas.DataFrame形式)：匿名化データセット ata_data(pandas.DataFrame形式)：攻撃対象データセット att(list[string]形式)：ユークリッド距離を測定する属性のリスト n(int形式)：各匿名化レコードについてリストアップする候補レコード件数	Dc(list[list[int]]形式)：各匿名化レコードに対応する攻撃対象データセットのn件のレコードのインデックスリスト Dc_yuclid(list[list[float]])：匿名化データと攻撃データのユークリッド距離リスト	1件の匿名化レコードに対して、ユークリッド距離が近い攻撃対象データセットのレコードをn件リストアップする
	Attack_id(Dc, Dc_yuclid, duplication)	Dc(list[list[int]]形式)：匿名化データに対応する攻撃データn件のインデックスリスト Dc_yuclid(list[list[float]])：匿名化データと攻撃データのユークリッド距離リスト Duplication(bool形式)：Trueなら重複を許してユークリッド距離最短のレコード同士をマッチングさせ、Falseなら重複を許さずにユークリッド距離が短い順に各レコードのマッチングを行う	各匿名化レコードに対応する攻撃レコード1件のインデックスリスト	Neighbor_attで作成した1:nのリストDcを1:1に絞り込む
exe_attack_code.py			匿名化データセットの各レコードと攻撃対象データセットのレコードのインデックスリスト	再識別攻撃のサンプル実行コード

演習課題

サンプルコード

※n=3のとき

Dc

```
[
    [4, 119, 53],
    [72, 10, 124],
    [30, 148, 51],
    ・・・
]
```

Dc_yuclid

```
[
    [0.2, 0.24, 0.4],
    [0.01, 0.06, 0.14],
    [0.0, 0.1, 1.0],
    ・・・
]
```

attack_id()

correspond

```
[
    4,
    10,
    30,
    ・・・
]
```

匿名データ0番は
攻撃データの4番、119番、53番と
対応する可能性がある・・・という要領

Dcと対応。
例えば、匿名データ1番と攻撃データ
72番のユークリッド距離は0.01となる

匿名データ0番と攻撃データ4番,
匿名データ1番と攻撃データ10番,
匿名データ2番と攻撃データ30番
が対応とする

演習課題

注意点（制限）

・属性値は元データセットに存在する値とすること

　ー一般化の実施は可能だが、最終的な属性値は第一階層の値とすること

　（e.g.）”東京”を”関東”などに一般化してもよいが、最終的に提出するデータは例えばランダムで埼玉、東京に置き換える

・レコード削除を行う際は、当該レコードの全属性値を欠損（空白）とすること

・その他質問がある場合はslack？