論述 》統計 · 調查

美國人口普查之差分隱私技術

近年美國普查局面對數據重建與重新識別的威脅，逐步引入更高階的隱私保護技術，不僅具有數學可證明性，還能根據需求靈活設定保護強度，可有效防止統計數據被還原為個人資料。2020 年美國人口普查首次應用差分隱私（Differential Privacy）技術，在統計資料中添加雜訊（Noise），藉此平衡數據的準確性與隱私性，並確保公開統計數據兼顧安全與實用性。

謝博行（行政院主計總處國勢普查處專員）

壹、前言

人口普查釋出之統計資料多樣且複雜，如果發布的範圍很小且多數人具有相似特徵，當該區域某個人具有獨特特徵，很容易猜測出當事人的身分。隨電腦算力提升及各類資料庫數量快速增加，美國普查局（U.S. Census Bureau）發現，即使在較大地理區域中，統計資料越來越容易受到重建資料庫（Database Reconstruction）和重新識別（Reidentification）的威脅，人口特徵被揭露的風險比以往更高。

美國普查局曾使用 2010 年人口普查統計結果表（業經避免資料揭露技術處理）進行一項實驗，結果成功重建逾 3 億筆個人資料；隨後，美國普查局將這些紀錄與 4 個商業資料庫進行比對，識別出居民的年齡、性別、種族、族裔和地理位置等資訊，再比對人口普查原始資料，一致率達 91.8%（下頁圖 1），一旦重建的資料被公開，將違反美國人口普查的保密承諾。

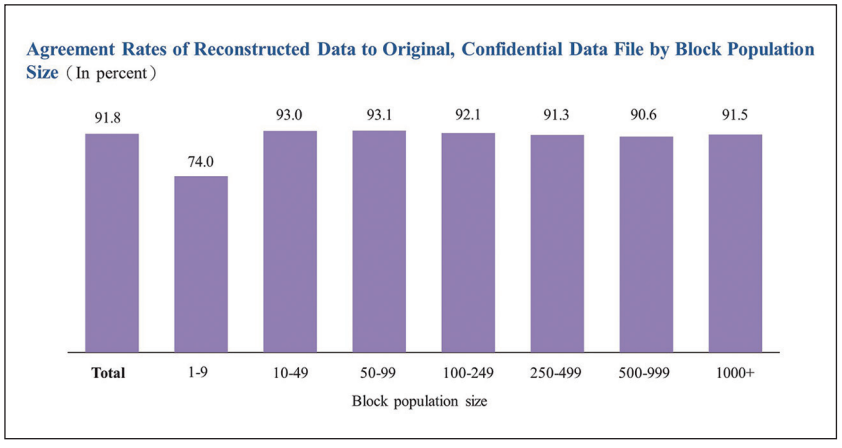
這項漏洞主因乃 2010 年美國人口普查所採用的資料保護方法－資料交換（Data Swapping），其目的在於保護小群體中可能被識別的個人資料，但無法阻擋資料庫重建和重新識別等類型的攻擊；美國普查局資料管理執行政策委員會（Data Stewardship Executive Policy Committee）因而對其揭露避免（Disclosure Avoidance）技術加以改進－「差分隱私」保護機制應運而生。

貳、差分隱私概念

「ε-差分隱私（ε-Differential Privacy）[1]」的概念係由 Dwork 等人在 2006 年提出，其核心觀念是「當某筆資料被加入或移除後，外界取得的資訊差異不大，從而防止單一個人的資訊被識別或推斷出來；如果資訊差異的大小得以控制，就能評估資料被揭露的最大風險」，差分隱私不同於傳統的揭露避免方法，以數學定義統計資料集的隱私損失，其雜訊添加技術提供「可量化」且「可證明」的保證，比原有的資料保護方式更透明，更能保持統計數據的準確性和實用性。

差分隱私實際的做法是在統計表中添加雜訊，也就是隨機在每個查詢結果加或減一個數值；藉由控制雜訊的大小，限縮外界可獲得的資訊，避免已發布統計表被任意組合而推斷出特定個人或家庭特徵。就像電視螢幕上的圖像，實際上是由數百萬個像素組成，添加雜訊類似對「像素」進行微小更改，當影像縮小時仍可完整呈現整體圖像，惟正確識別單一像素的風險大幅降低。美國目前採用的零集中差分隱私（Zero-Concentrated Differential Privacy）是傳統差分隱私的一種強化版本，能更加靈活控制隱私損失，並搭配離散型高斯分布（Discrete Gaussian Distribution）作爲添加雜訊的機制，以確保人數統計結果爲整數。

圖 1　資料庫重建結果與人口普查資料比對一致率



資料來源：美國普查局。

參、隱私保護的演算法架構

爲保護 2020 年人口普查受訪者的回復資料，美國普查局建構揭露避免系統（Disclosure Avoidance System），以差分隱私概念爲基礎，針對不同統計產品特性開發 TDA（TopDown Algorithm）、SafeTab 及 PHSafe 三種演算法，以下分別摘述其內涵：

一、TDA 演算法

TDA 演算法主要用於基礎人口統計表，目標是「小地區資料能受到保護，大地區資料則要求準確」。其處理流程分爲 5 個步驟，首先將個別資料檔（Microdata）與地理參照檔（Geographic Reference File）合併產生製表檔案；第 2 個步驟再將各地理級別（從國家、州，到普查街廓）與所有人口特徵變數類別交叉，資料轉換

論述 》統計 · 調查

成分組統計結果（Conversion to Histogram）。

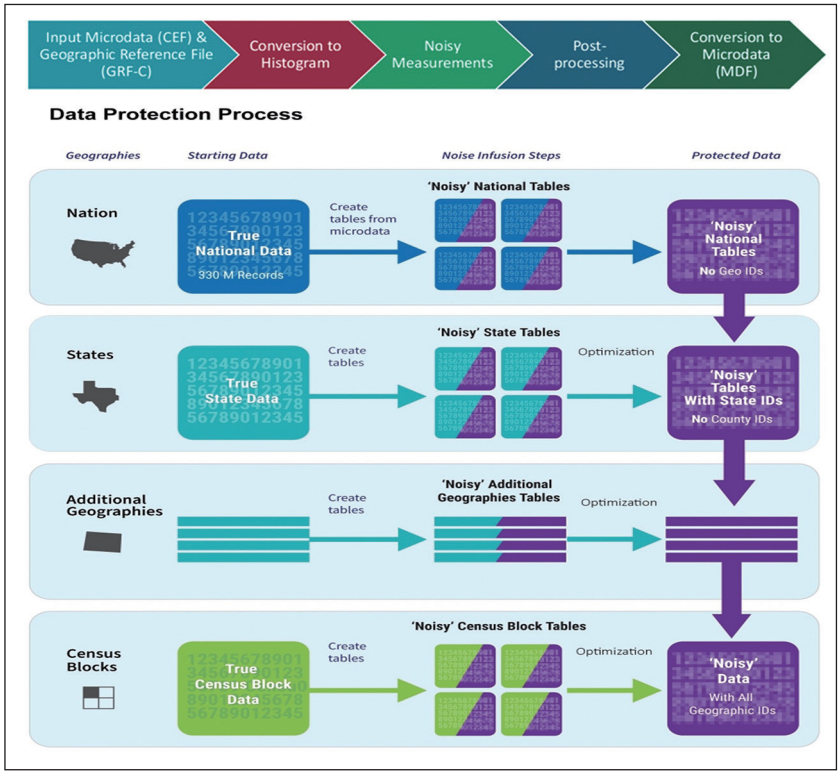
第 3 個步驟在分組統計結果的每個單元格添加隨機雜訊（Noisy Measurements）（圖 2），雜訊量與其原本數值大小無關，並且獨立添加其中，所以數值小的單元格有可能添加較大雜訊量，抑或部分單元格未添加任何雜訊（數值保持不變），並因而導致表格內數據出現邏輯上的錯誤（例如：某些統計人口數爲負值，或者州內各郡人口數加總後不等於該州總人口數）。

第 4 個步驟「後處理（Post-processing）」則用以調整這些不合理現象，包括「不變量（Invariants）」和「額外限制條件（Additional Constraints）」。美國人口普查結果的各州人口統計被用來重新分配美國眾議院席位，因此「不變量」係確保某些數值在雜訊添加後不會變動，如「各州的總人口數」、「各普查街廓的住宅單位總數」與「各普查街廓各類型使用中的集體住所設施數量」等；「額外限制條件」則以強制調整數值的方式處理資料不合理現象，如限定人口和住宅數量不爲負值、各統計表的行（列）單元格加總應與其邊際值相同等。

TDA 演算法是由上至下（Top-Down）方式處理，從國家層級開始，各項資料確定後，再處理州層級，州層級各項總和應與國家層級一致，越大區域的統計結果越準確，不會因數值加總造成雜訊累積，經過「後處理」方可產生受保護且合理的統計表。

最後一個步驟主要運用受保護統計表推估各項特徵機率，生成「隱私保護微數據檔案（Privacy-Protected Microdata File）」，模擬出的微數據可用於編表系統，產製各項統計結果並提供外界應用。

圖 2　TDA 演算法



資料來源：美國普查局。

二、SafeTab 演算法

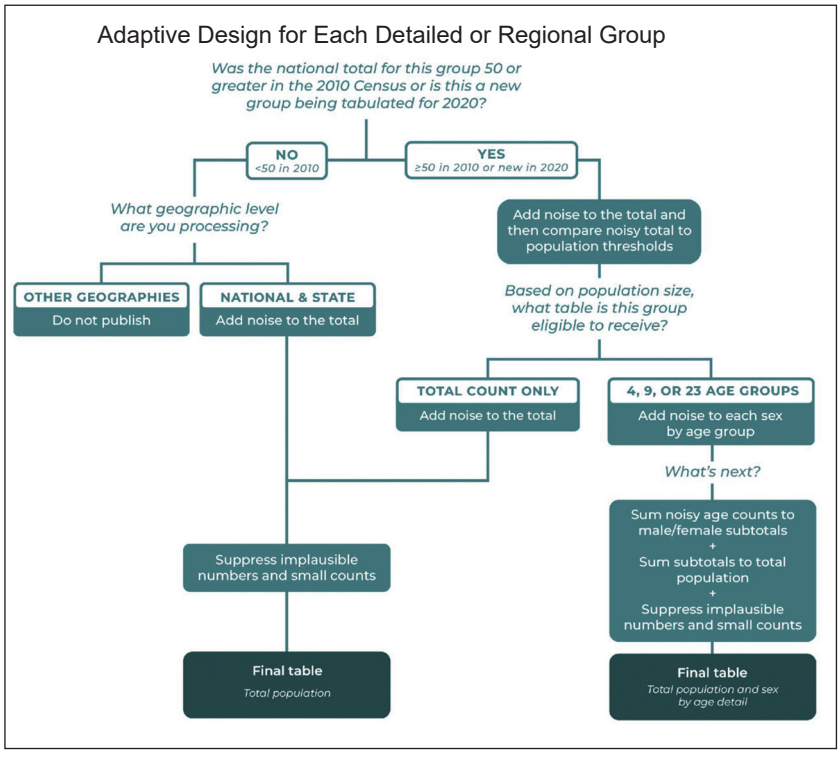
SafeTab 演算法主要用於

提供全國種族、族裔、美國原住民及阿拉斯加原住民部落等詳細數據，由於部分群體的人數相對較少，發布相關統計數據並兼顧個人隱私成爲一項挑戰。

SafeTab 演算法的重點爲「自動調整設計」，在考量不同群體的人數規模後，發布不同細緻程度的統計表，盡可能在機密保護前提下提供可用的人口統計，以滿足資料使用者的需求。以「詳細人口與住宅特徵檔案（Detailed Demographic and Housing Characteristics File）A」的處理爲例（圖 3），SafeTab 先檢查每個群體是否「在 2010 年人口普查中全國人口數少於 50」，如果是，則僅發布全國和州層級添加雜訊的統計結果，不會產生性別或年齡別數據；反之，如果該群體「在 2010 年人口普查中全國人口數爲 50 以上，或爲 2020 年普查蒐集到的新群體」，則先計算加入雜訊後的總人口數，再依預先設定之人口門檻值決定統計表內容（總人口數愈多，所發布之統計表愈詳細）。然後，對於可發布「性別按年齡組分」資料的群體，在每個性別與年齡組別的統計值獨立地添加雜訊，分別加總後得到男性及女性人口數，性別資料再相加計算總人口數。最後檢視統計表內容，對於人數較小或不合理結果予以遮蔽，產生對外公開資料。

由於 SafeTab 演算法會在不同地理區域和人口群體中獨立重複添加雜訊，可能導致統計表之間數據不一致，或資料加總後雜訊被放大；因此，美國普查局建議使用者不宜自行加總數據，而是使用官方已發布的現成資料，以免獲得不精確的結果。

圖 3　SafeTab 自動調整設計



資料來源：美國普查局。

三、PHSafe 演算法

PHSafe 演算法主要用於個人與住宅數據結合的統計表，例如平均家戶規模、家戶人口

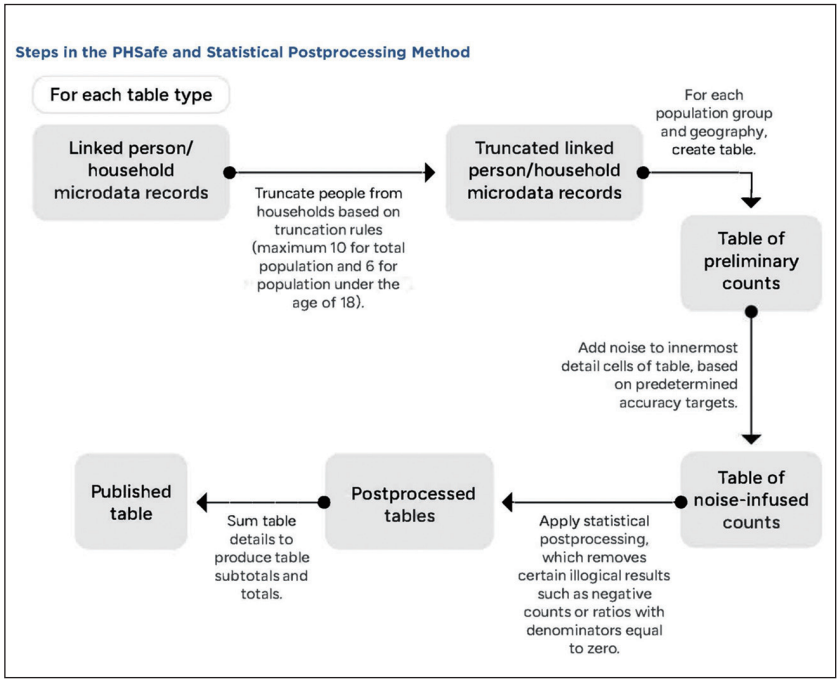
論述 》統計 · 調查

之家戶型態等統計結果。當「個人檔案」和「住宅單位檔案」合併後，便能提供家戶內所有成員的資訊，包括家戶成員之間的關係；但也因此難以隱藏個人數據對其他成員的影響，進而增加保護這些數據的難度，使整合後統計結果的揭露風險比單獨發布其中一項資料更高，所以僅能提供「州」以上地理區域的特定表格。

爲保護家戶特徵，PHSafe 演算法主要係透過隨機移除部分家戶成員的截斷（Truncation）機制，來限制家戶人口數（圖 4）。截斷規則係根據家戶「總人口數」（上限爲 10 人）或「未滿 18 歲人口數」（上限爲 6 人）採取不同的步驟，首先每個家戶成員隨機分配到一個索引號，並依據該索引號進行排序，針對統計標的爲「家戶總人口」的結果表，第 10 位以後的成員會被移除；而對於統計標的爲「未滿 18 歲人口」的結果表，則先排除年滿 18 歲的成員，再就未滿 18 歲者按照索引號從低到高排序，移除第 6 位以後的成員。

接下來就截斷處理後的資料進行特徵別初步統計，並對計算結果添加雜訊；最後爲了減少雜訊造成的不合理情形，再以貝氏方法（Bayesian Approach）進行「後處理」。公布的統計結果是基於截斷後的人口數據，而非原始的人口數據，可確保數據符合邏輯，且受到充分保護。

圖 4　PHSafe 和後處理流程



資料來源：美國普查局。

肆、結語

隨著電腦算力提升，大幅改善統計資料處理效率，使政府更能強化數據治理的應用。爲滿足各界對資料的殷切需求，公開的統計資料越來越多，以往認爲只要不揭露特殊少數人口群體的統計結果，就能有效防止個別資料被辨識；現今面對重建資料庫與重新識別的威脅，部分人口群體已成爲新的攻擊目標，而傳統方法的保護效果卻很有限。

美國普查局爲確保發布的資料具有一定程度的保護力，

對於 2020 年人口普查提供的各種統計產品，開發 TDA、SafeTab 及 PHSafe 等演算法，以因應不同層次的保護需求。每一種演算法各有其特點與應用場合，TDA 採用地區層級與不變量框架，使範圍越大的地區有越精準資料；SafeTab 導入自動調整設計，避免發布過度精細的資料；PHSafe 採用截斷技巧，降低住宅單位與個人檔案串聯所增加的風險。

差分隱私技術的使用並非一成不變，唯有深入了解資料特性與使用情境，才能兼顧資料保護及其可用性，美國隱私保護技術的發展與應用值得我國借鏡。

註釋

1. 就給定的正實數 ε 而言，倘若一個隨機演算法 M，對於所有兩個相鄰資料集 x 與 x'（僅差一筆資料），可能的查詢結果 E 均滿足 $P(M(x) \in E) \leq e^{\varepsilon} P(M(x') \in E)$ ，則 M 稱爲 ε- 差分隱私（ε - differential privacy）；這個條件確保查詢結果受單筆資料變動的影響控制在 e^{ε} 倍以內，ε 用來調整隱私保護的強度，數值越小表示保護力越強。

參考文獻

1. 陳艷秋（2022），美國人口普查避免個別資料揭露方法之變革，主計月刊，802 期，74-79 頁。
2. U.S. Census Bureau. (2024a), Disclosure Avoidance and the Supplemental Demographic and Housing Characteristics File (S-DHC): How PHSafe Works, U.S. Department of Commerce.
3. U.S. Census Bureau (2024b), U.S. Census Bureau Workshop on Using 2020 Census Data: New guidance and resources for assessing the fitness-for-use of differential privacy-adjusted Census data.
4. U.S. Census Bureau (2023a), Why the Census Bureau Chose Differential Privacy, U.S. Department of Commerce.
5. U.S. Census Bureau. (2023b), Disclosure Avoidance and the 2020 Census: How the TopDown Algorithm Works, U.S. Department of Commerce.
6. U.S. Census Bureau. (2023c), Disclosure Avoidance Methods for the Detailed Demographic and Housing Characteristics File A (Detailed DHC-A): How SafeTab-P Works, U.S. Department of Commerce.
7. U.S. Census Bureau (2021a), Understanding the 2020 Census Disclosure Avoidance System: Differential Privacy 201 and the TopDown Algorithm.
8. U.S. Census Bureau (2021b), Disclosure Avoidance for the 2020 Census: An Introduction, U.S. Department of Commerce.
9. Dwork, C., McSherry, F., Nissim, K., Smith, A. (2006), Calibrating Noise to Sensitivity in Private Data Analysis, Theory of Cryptography(TCC 2006), Springer.❖