Can swapping be differentially private?
A refreshment stirred, not shaken

James Bailie, Ruobin Gong and Xiao-Li Meng

January 19, 2023

For Internal Review Purposes Only – Please Do Not Circulate

Extended Abstract

This paper presents a formal privacy analysis of data swapping [Dalenius and Reiss, 1982, Fienberg and McIntyre, 2004]. Data swapping refers to a general class of statistical disclosure control (SDC) methods which interchanges the sensitive variables of a random subset of records. As the main class of methods for confidentiality protection, data swapping is central to the design of the 1990, 2000 and 2010 U.S. Decennial Census disclosure avoidance systems (DAS) [McKenna, 2018].

The goals of this paper are three-fold. First, we develop a theoretical understanding of the privacy guarantees that swapping affords. In order to conduct a rigorous privacy analysis, we focus on one type of swapping algorithm which was constructed to be congenial with formal privacy analysis while also aligning as much as possible with the public knowledge of the 2010 Census's swapping method. Like all swapping algorithms, the algorithm we examine has *invariants* (that is, statistics calculated from the confidential database which remain unchanged) and thus does not satisfy the classic notion of pure differential privacy (ϵ-DP). However, we show that the swapping algorithm can be examined through a rigorous, DP-like lens by explicitly acknowledging the nuances of its privacy guarantee.

Our second goal is to establish swapping methods on an equal and formal footing with other invariant-respecting DP mechanisms, in order to compare and elucidate their similarities and differences. Swapping mechanisms have received criticism since they have been shown theoretically to introduce bias into the published data [Drechsler and Reiter, 2010]. But the level of this bias depends on implementation parameters which are typically kept secret. Only with public transparency of the 2010 swapping algorithm – as enabled by a formal privacy analysis – can the extent and direction of this bias be quantified. This would provide practical considerations for the 2010 Census data, above and beyond what the current theoretical understanding can provide. With this formal privacy analysis serving as the theoretical support, a focal comparison we present in this

paper is between the 2010 Census DAS and the 2020 DAS. We prove that the swapping algorithm we examine, which mimics the 2010 Census DAS, satisfies a flavor of formal privacy called $(\boldsymbol{c}, \epsilon)$-DP that is similar to that of the 2020 DAS. Thus, this analysis illustrates the evolution over time of the privacy guarantees provided in the US Dicennial Censuses.

This paper continues an existing line of research [Bailie and Chien, 2019, Rinott et al., 2018] examining non-DP statistical disclosure control (SDC) techniques – which are typically regarded as ad-hoc – under the light of formal privacy. Hence our third and broadest goal is to demonstrate how seemingly ad-hoc SDC techniques can (and should) be framed and analyzed formally and rigorously. Since both DP and traditional SDC each have their own unique advantages, combining these two conflicting fields bestows opportunities to reap the best of both worlds. On the one hand, DP supplies a formal, mathematical description of the level and substance of privacy provided by a confidentiality method. DP also allows for complete transparency of the method without any degradation of these privacy guarantees. Recasting swapping algorithms, or other SDC techniques, as formally private provides strong guarantees that the details of these algorithms – which have traditionally been kept secret – can safely be made public. This transparency is an important prerequisite for any valid statistical analysis of privacy-protected data [Gong, 2022]. For example, our work suggests that the details of the 2010 DAS can be published without privacy risk, allowing for the first time statisticians, economists and social scientists to appropriately account for the DAS in their analyses. On the other hand, swapping carries its own advantages, including facial validity which is important to data users, and which DP methods, such as the 2020 DAS, typically do not achieve [boyd and Sarathy, 2022, Hotz and Salvo, 2022, Ruggles et al., 2019].

The swapping algorithm we examine can be briefly summarised as follows. The variable set is partitioned into two non-empty categories: the *swapping variables* $\boldsymbol{V}_{\text{Swap}}$ and the *holding variables* $\boldsymbol{V}_{\text{Hold}}$. Each record is independently selected with probability given by a parameter p called the *swap rate*. The swapping variables $\boldsymbol{V}_{\text{Swap}}$ of the selected records are then randomly shuffled. More specifically, a derangement σ (i.e. a permutation with no fixed points) over the selected records' indices is sampled uniformly at random. The new $\boldsymbol{V}_{\text{Swap}}$ of a selected record i is given by the $\boldsymbol{V}_{\text{Swap}}$ of the $\sigma(i)$-th record.

There is one further pertinent detail. Swapping is restricted to records which share the same values on a subset of the holding variables, called the *matching variables* $\boldsymbol{V}_{\text{Match}}$. That is, the matching variables $\boldsymbol{V}_{\text{Match}}$ define strata so that swapping is restricted within these strata. Thus, the above procedure is repeated independently within each category of $\boldsymbol{V}_{\text{Match}}$.

Like the 2020 DAS, swapping maintains some statistics as invariant. Specifically, there are two contingency tables which remain unchanged by swapping: 1) $\boldsymbol{V}_{\text{Match}} \times \boldsymbol{V}_{\text{Swap}}$: the cross-classification of the matching variables by the swapping variables; and 2) $\boldsymbol{V}_{\text{Match}} \times (\boldsymbol{V}_{\text{Hold}} - \boldsymbol{V}_{\text{Match}})$: the cross-classification of the matching variables by the other (i.e. non-matching) holding variables. We denote these two invariant tables by $\boldsymbol{c}_{\text{Swap}}$.

The following Theorem is a simplification of our paper's main contribution. In this abstract, we

focus on the typical case of $0 < p \leq 0.5$ for simplicity, but there are findings for general p in the paper.

Theorem. *The above swapping mechanism is* $(\boldsymbol{c}_{Swap}, \epsilon)$*-differentially private, where*

$$\epsilon = 0.5 \ln \left(\Theta(b^2)(1-p)^2 p^{-2} \right),$$

and b *is the number of records in the largest category of* $\boldsymbol{V}_{Match}$*.*

In intuitive terms, this Theorem is saying that the swapping algorithm we consider satisfies ϵ-differential privacy, when conditioning on the swapping invariants $\boldsymbol{c}_{\text{Swap}}$. What does this mean? In classic DP, ϵ can informally be viewed as a measure of the difference in output between any two input datasets. By conditioning on the swapping invariants, we restrict our consideration to input datasets which have the same values on $\boldsymbol{c}_{\text{Swap}}$. More exactly, the data universe $\mathcal{D}$ is now a function of the confidential dataset $\boldsymbol{X}^*$:

$$\mathcal{D}(\boldsymbol{X}^*) = \{\boldsymbol{X} : \boldsymbol{c}_{\text{Swap}}(\boldsymbol{X}) = \boldsymbol{c}_{\text{Swap}}(\boldsymbol{X}^*)\},$$

and we measure the difference in output between any two $\boldsymbol{X}, \boldsymbol{X'} \in \mathcal{D}(\boldsymbol{X}^*)$. In order to maintain the robustness property of DP, we must analyse every data universe $\mathcal{D}(\boldsymbol{X}^*)$, over all the possible confidential datasets $\boldsymbol{X}^*$. While classic DP compares all neighboring datasets in $\bigcup_{\boldsymbol{X}^*} \mathcal{D}(\boldsymbol{X}^*)$ simultaneously, invariant-respecting DP analyzes each $\mathcal{D}(\boldsymbol{X}^*)$ separately. Rather than being ad-hoc, this restriction is a necessary precondition for defining formal privacy with invariants. If we were required to compare datasets across $\mathcal{D}(\boldsymbol{X}^*)$, then we could no longer release invariants exactly.

Separating the data universe into classes $\mathcal{D}(\boldsymbol{X}^*)$ raises one further complication: DP is traditionally defined by comparing neighboring datasets in $\mathcal{D}$, but there may not be any pairs of neighboring datasets in $\mathcal{D}(\boldsymbol{X}^*)$. We therefore consider an equivalent definition of DP using distance metrics d on datasets (such as the Hamming distance) in place of neighbours.

Definition. A mechanism M satisfies $(\boldsymbol{c}, \epsilon)$*-differential privacy* if, for all $\boldsymbol{X}^*$, all possible outputs o and all datasets $\boldsymbol{X}, \boldsymbol{X'} \in \mathcal{D}(\boldsymbol{X'}) = \{\boldsymbol{X} : \boldsymbol{c}(\boldsymbol{X}) = \boldsymbol{c}(\boldsymbol{X}^*)\}$, the following inequality is satisfied:

$$\Pr(M(\boldsymbol{X}) = o) \leq \exp\left(\epsilon d(\boldsymbol{X}, \boldsymbol{X'})\right) \Pr(M(\boldsymbol{X'}) = o).$$

This discussion highlights some of the important nuances underlying any formal privacy definition. Other nuances, as discussed in the paper, include the choice of divergence on the output space: pure DP uses the multiplicative distance, while other flavors of DP such as (ϵ, δ)-DP and ρ-zero concentrated DP (zCDP) use different divergences. Revealing these nuances allows the definition of differential privacy to be reinterpreted as a broad standard that can be used to investigate and compare a wide range of confidentiality methods. Importantly, many existing implementations of differential privacy rely on convenient interpretations of these nuances. (For example, the 2020 US

Census's use of invariants [Abowd et al., 2022], or the use of neighbouring datasets which differ by only a single interaction, rather than an individaul's entire collection of interactions [Kenthapadi and Tran, 2018].) By recasting DP in a light which illuminates these nuances, we hope to improve the clarity and rigor in assessments of DP implementations.

As part of our comparative analysis of the 2010 DAS and the 2020 DAS, we show in the paper that the formal privacy of the 2020 DAS requires relaxations to both (ϵ, δ)-DP (or zCDP) and to allow for invariants. Thus the swapping mechanism we examine, which is designed to be similar to the 2010 Census DAS, satisfies a flavor of DP which is similar to that of the 2020 DAS. The within-system privacy evaluation (i.e. the privacy budget ϵ) is similar in magnitude between the 2020 Census Redistricting Data (P.L. 94-171) Summary File and the entire 2010 Census data releases as a whole (assuming a 2-4% swap rate as purported by boyd and Sarathy [2022]). However, the 2020 Census data has not been fully released yet, which means the 2020 privacy budget will increase further. Additionally, the privacy budgets for 2010 and 2020 cannot be compared directly, since the invariants of each Census are different and thus they satisfy different flavors of DP. This reflects the broader point that privacy budgets are contextual and their interpretation depends on a range of factors, including the sensitivity of the collected microdata and of the published statistics as well as the exact privacy definition used.

Our formal assessment of swapping showcases the *relative* privacy guarantee it affords. It does not, as no disclosure mechanism can, offer *absolute* privacy guarantees. While a determination of swapping's privacy budget ϵ provides a within-system privacy evaluation, across-system evaluations – comparing the disclosure risk of ϵ-DP mechanisms with different sets of invariants – is inherently subjective and contextual. We leave this as an important topic for future research.

References

J. Abowd, R. Ashmead, R. Cumings-Menon, S. Garfinkel, M. Heineck, C. Heiss, R. Johns, D. Kifer, P. Leclerc, A. Machanavajjhala, B. Moran, W. Sexton, M. Spence, and P. Zhuravlev. The 2020 Census Disclosure Avoidance System TopDown Algorithm. *Harvard Data Science Review*, (Special Issue 2), June 2022. doi: 10.1162/99608f92.529e3cb9.

J. Bailie and C.-H. Chien. ABS Perturbation Methodology Through the Lens of Differential Privacy. In *Work Session on Statistical Data Confidentiality, UN Economic Commission for Europe*, page 13, Oct. 2019.

d. boyd and J. Sarathy. Differential Perspectives: Epistemic Disconnects Surrounding the U.S. Census Bureau's Use of Differential Privacy. *Harvard Data Science Review*, (Special Issue 2), June 2022. doi: 10.1162/99608f92.66882f0e.

T. Dalenius and S. P. Reiss. Data-swapping: A technique for disclosure control. *Journal of Statistical*

Planning and Inference, 6(1):73–85, Jan. 1982. ISSN 0378-3758. doi: 10.1016/0378-3758(82)90058-1.

J. Drechsler and J. P. Reiter. Sampling With Synthesis: A New Approach for Releasing Public Use Census Microdata. *Journal of the American Statistical Association*, 105(492):1347–1357, Dec. 2010. ISSN 0162-1459. doi: 10.1198/jasa.2010.ap09480.

S. Fienberg and J. McIntyre. Data Swapping: Variations on a Theme by Dalenius and Reiss. In *Privacy in Statistical Databases*, 2004. doi: 10.1007/978-3-540-25955-8_2.

R. Gong. Transparent Privacy is Principled Privacy. *Harvard Data Science Review*, (Special Issue 2), June 2022. doi: 10.1162/99608f92.b5d3faaa.

V. J. Hotz and J. Salvo. A chronicle of the application of differential privacy to the 2020 Census. *Harvard Data Science Review*, (Special Issue 2), June 2022. ISSN 2644-2353, 2688-8513. doi: 10.1162/99608f92.ff891fe5.

K. Kenthapadi and T. T. L. Tran. PriPeARL: A Framework for Privacy-Preserving Analytics and Reporting at LinkedIn. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, pages 2183–2191, New York, NY, USA, Oct. 2018. Association for Computing Machinery. ISBN 978-1-4503-6014-2. doi: 10.1145/3269206.3272031.

L. McKenna. Disclosure avoidance techniques used for the 1970 through 2010 Decennial Censuses of Population and Housing. Working paper, The Research and Methodology Directorate - US Census Bureau, Nov. 2018. URL https://www.census.gov/content/dam/Census/library/working-papers/2018/adrm/Disclosure%20Avoidance%20for%20the%201970-2010%20Censuses.pdf.

Y. Rinott, C. M. O'Keefe, N. Shlomo, and C. Skinner. Confidentiality and Differential Privacy in the Dissemination of Frequency Tables. *Statistical Science*, 33(3):358–385, Aug. 2018. ISSN 0883-4237. doi: 10.1214/17-STS641.

S. Ruggles, C. Fitch, D. Magnuson, and J. Schroeder. Differential privacy and Census data: Implications for social and economic research. *AEA Papers and Proceedings*, 109:403–408, May 2019. ISSN 2574-0768. doi: 10.1257/pandp.20191107.