Supplementary Methods Appendix to Differential privacy in the 2020 US census: what will it do? Quantifying the accuracy/privacy tradeoff: Design and validation of Empirical Privacy Loss (EPL) metric

Abraham D. Flaxman and Samantha Petti, 2020-03-25

A randomized algorithm $\mathcal{A}$ is ϵ-DP if, for each event E, for any pair of datasets D and D' that are the same everywhere except for on one person's data,

$$\Pr[\mathcal{A}(D) = E] \leq \exp(\epsilon)\Pr[\mathcal{A}(D') = E].$$

This bound is typically proven by logical deduction, and for complex DP algorithms, the proof often relies on the Sequential Composition Theorem, which states that information derived by combining the output of an ϵ_1-DP algorithm and an ϵ_2-DP algorithm is at most $(\epsilon_1 + \epsilon_2)$-DP. This theorem is an inequality, however, and the inequality might have room for improvement.

As mentioned in the main text above, it is possible to empirically quantify privacy loss and thereby see if the inequality of the sequential composition theorem is not tight. A brute force approach to do this is to search over databases D and D' that differ on one row to find the event E with the largest ratio of probabilities. However, this search is too computationally intensive to be feasible for all but the simplest DP algorithms.

For algorithms that produce DP counts of multiple subpopulations, such as TopDown, we propose using the distribution of the residual difference between the precise count and the DP count as a proxy for the distribution produced by the brute force approach. We first assume that the residual difference of the DP count minus the precise count is identically distributed for queries across similar areas (such as voting-age population across all enumeration districts). Then, instead of focusing on only the histogram counts containing the individual who has changed, we use the residuals for all areal units to estimate the probability of the event we are after as

$$\Pr\left[\text{error}_j = k\right] \approx \left(\sum_{j'=1}^{C} \mathbf{1}\left[\left\{\text{error}_{j'} = k\right\}\right]\right)/C =: \hat{p}_k,$$

where error_j is the residual difference of DP counts returned by TopDown minus the precise count for that same quantity in the 1940 census, and the $\text{error}_{j'}$ are residuals for C other queries assumed to be exchangeable.

After making this assumption, we define the empirical privacy loss for any residual x, which we denote by $\text{EPL}(x)$, in terms of an approximation of probability distribution of the residuals (DP count minus precise count at a selected level of the geographic hierarchy), which we denote $p^{\text{KDE}}(x)$, using Gaussian kernel density estimation with a bandwidth of 0.1 to smooth $\hat{p}_k$, and compare the log-ratio inspired by the definition of ϵ-DP algorithms:

$$\text{EPL}(x) = \log\left(\frac{p^{\text{KDE}}(x)}{p^{\text{KDE}}(x+1)}\right).$$

The empirical privacy loss (denoted by EPL) is then the maximum of the absolute value of $\text{EPL}(x)$ for any x:

$$\text{EPL} = \max_{x \in (-\infty,\infty)} \{\text{abs}(\text{EPL}(x))\}$$

For a simple demonstration of how EPL functions, we now present an example of EPL when applied to a parallel geometric mechanism. We begin with the precise counts of the total population for enumeration districts in Washington State from the 1940 census and then produce geometric DP (GDP) versions of these counts by adding variation drawn independently at random from a symmetric geometric distribution with parameter $\epsilon = 0.05$. In the notation above, there are $C = 2{,}663$ such queries, and the residual for each $j = 1, \ldots, C$ is independent and identically distributed with two-tailed geometric distribution of $\text{error}_j \sim G(\epsilon/2)$. (Figure SA1)

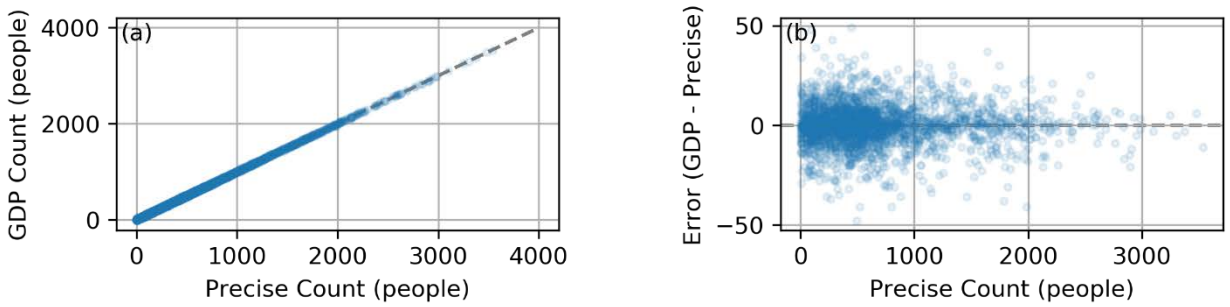

Figure SA1: Panel (a) shows a scatter plot of the precise total count of people versus a corresponding DP count produced by the parallel application of the geometric mechanism with $\epsilon = 0.25$ for each enumeration district in Washington State, based on 1940 census data. Panel (b) shows a scatter plot of the residual difference of the DP count minus the precise count as a function of the precise count. Note that, unlike TopDown, the geometric mechanism does not guarantee DP counts are non-negative.

We then calculate the EPL as described above. First, we use the residual differences between the DP and precise counts as input to a Gaussian kernel-density estimator (with bandwidth 0.1) to estimate $p^{\text{KDE}}(x)$. Then we search over x ranging from the 5$^{\text{th}}$ to 95$^{\text{th}}$ percentiles of the residual distribution to find the maximum absolute value of the log of the ratio of $p^{\text{KDE}}(x)/p^{\text{KDE}}(x+1)$ (Figure SA2). In this example application to the Geometric Mechanism, we find EPL of 0.043, which is not too far from the theoretical value of $\epsilon = 0.05$.

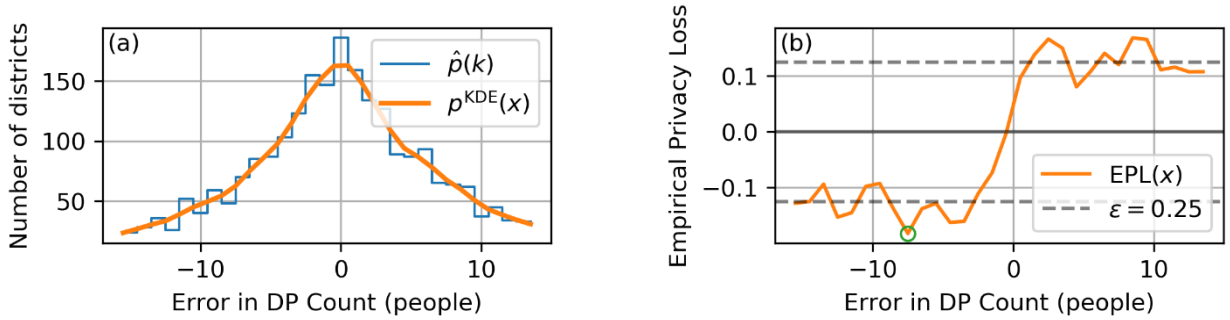

Figure SA2: Panel (a) shows the direct and KDE-smoothed estimates of the density of residuals for the geometric mechanism, and panel (b) shows the corresponding EPL(x) and how it yields a maximum absolute value of 0.18, which is similar in value to the theoretical value of $\epsilon = 0.25$.

The detail-oriented reader may have noted that the choice of bandwidth of 0.1 and limits for maximization of 5th to 95th percentiles were somewhat arbitrary, but we hope they would also be reassured that the EPL value is close to ϵ for this simple mechanism. We validated our approach by repeating this calculation for a range of epsilon values, and performed a one-way sensitivity analysis to see if changing the bandwidth or limits for maximization would substantially change the ability of the method to recover an EPL value matching epsilon for the Geometric Mechanism. (Table SA1.)

[TABLE SA1 HERE]

EPL is capable of producing evidence that the bounds in the sequential composition theorem are not tight, as well. As a minimal example of how this can happen, we add a little complexity to our simple demonstration example, by including a measurement of the total count in each enumeration district with $\epsilon_1 = 0.25$ and the measurements of the stratified counts in each enumeration district with $\epsilon_2 = 0.75$. Inspired by the optimization step of TopDown, we then use optimization to find consistent values for the detailed counts that minimize the weighted sum of the absolute differences between the measured values and the optimized values. This is $(\epsilon_1 + \epsilon_2)$-DP because of sequential composition, but the EPL is substantially smaller than this sum, with a value of 0.54. (Figure SA3)

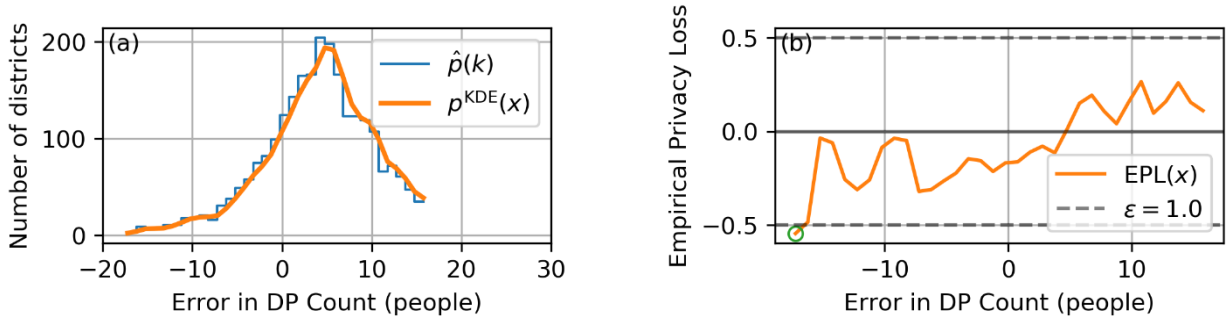

Figure SA3: Panel (a) shows the direct and KDE-smoothed estimates of the density of residuals for a DP mechanism that combines geometric, and panel (b) shows the corresponding EPL(x) and how it yields a maximum absolute value of 0.54, which is smaller than the value of $\epsilon = 1.0$ proven using the sequential composition theorem.

TopDown includes invariants, which are a feature that goes beyond the traditional formulation of epsilon-DP. As a final example, we consider how an invariant can produce an EPL that is *greater* than the privacy loss budget epsilon that is formally proven for the algorithm without invariants. To achieve this, we now shift the example of EPL of stratified counts (stratified by voting age, race, and ethnicity) with geometrically distributed variation added to achieve DP with $\epsilon = 0.25$, but we combine this with an optimization step that includes invariants of the total population count at the county level. In this case, the EPL of the optimized count (stratified and at the enumeration district level) is larger the theoretical ϵ , with EPL=2.3. This shows how the EPL construct is capable of quantifying the degree to which invariants compromise differential privacy. (Figure SA4)

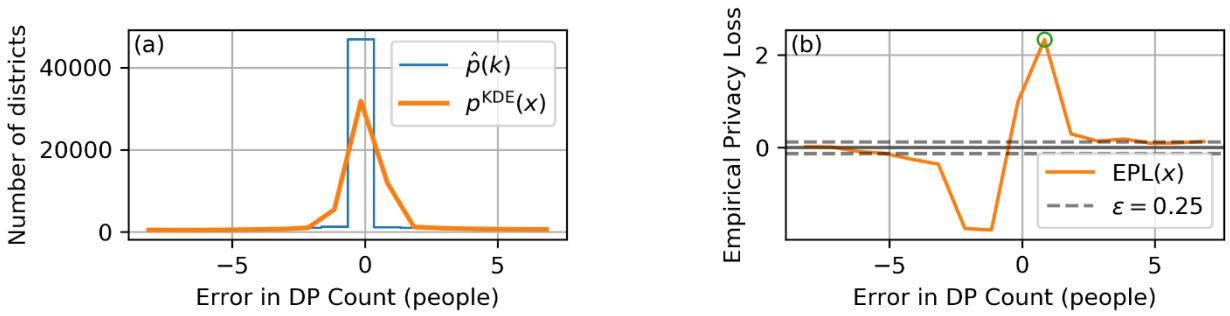

Figure SA4: Panel (a) shows the direct and KDE-smoothed estimates of the density of residuals for a mechanism that includes an invariant and therefore is not formally DP, and panel (b) shows the corresponding EPL(x) and how it yields a maximum absolute value of 2.3, which is substantially larger than the $\epsilon = 0.25$ used in the geometric DP portion of the mechanism.

Table SA1: Validation of empirical privacy loss construct comparing EPL values to geometric DP counts for a range of (a) epsilon values for bandwidth=0.1 and search percentile=95; (b) KDE bandwidth values for epsilon=0.25 and search percentile=95; and (c) search range percentiles for epsilon=0.25 and bandwidth=0.1.

(a) epsilon value in Geometric DP	EPL mean	EPL Lower Bound - 2.5th Percentile	EPL Upper Bound - 97.5th Percentile
0.0010	0.0010	0.0008	0.0013
0.0050	0.0048	0.0039	0.0068
0.0100	0.0099	0.0076	0.0130
0.0500	0.0490	0.0390	0.0673
0.1000	0.0980	0.0752	0.1262
0.1500	0.1475	0.1181	0.1941
0.2000	0.1988	0.1521	0.2639
0.2500	0.2429	0.1853	0.3493
0.3000	0.2824	0.2228	0.3806
0.3500	0.3252	0.2651	0.4116
0.4000	0.3482	0.2717	0.4360
0.4500	0.3827	0.3140	0.4807
0.5000	0.4052	0.3434	0.5195

(b) KDE band-width value	EPL mean	EPL Lower Bound - 2.5th Percentile	EPL Upper Bound - 97.5th Percentile
0.0010	0.2385	0.1872	0.3257
0.0050	0.2327	0.1853	0.3130
0.0100	0.2396	0.1837	0.3410
0.0500	0.2364	0.1911	0.3246
0.1000	0.2410	0.1858	0.3195
0.5000	0.2366	0.1842	0.3146
1.0000	0.2431	0.1946	0.3251
5.0000	0.2429	0.1853	0.3493

(c) EPL Search Range Percentile	EPL mean	EPL Lower Bound - 2.5th Percentile	EPL Upper Bound - 97.5th Percentile
75.0000	0.2385	0.1872	0.3257
85.0000	0.2327	0.1853	0.3130
95.0000	0.2396	0.1837	0.3410
99.0000	0.2364	0.1911	0.3246
99.9000	0.2410	0.1858	0.3195