Private Selection with Heterogeneous Sensitivities

Daniela Antonova
Apple

Allegra Laro
Apple

Audra McMillan
Apple

Lorenz Wolf[*]
University College London
lorenz.wolf.22@ucl.ac.uk

Abstract

Differentially private (DP) selection is the problem of selecting a high scoring candidate from a finite candidate pool where the score of each candidate is a function of a sensitive dataset. This problem arises naturally in a variety of contexts including model selection, hypothesis testing, and as a subroutine in many differentially private algorithms. Classical work in private selection assumes that the "goodness" of all selection candidates is equally sensitive to changing the data of a single individual. However, in many settings this assumption is false. Intuitively, utilising the fact that some candidate scores are less sensitive should allow for more accurate differentially private selection mechanisms. In this work we theoretically and empirically compare DP selection algorithms that account for heterogeneity with standard DP selection algorithms that assume homogeneous sensitivities of the candidate scores. While the selection algorithms that account for heterogeneity can add less noise to the overall computation, we find that, surprisingly, this does not always result in a higher utility algorithm. Our theoretical analysis and experimental results shed light on scenarios when heterogeneous sensitivities can be exploited.

1 Introduction

Differentially private (DP) selection is a fundamental task that arises naturally in a variety of contexts including model selection, hypothesis testing and as a subroutine in many algorithms. Given a set of candidates $\mathcal{A}$, a score function $q : \mathcal{A} \times \chi^n \to \mathbb{R}$, and a database $D \in \chi^n$, a differentially private selection algorithm aims to output the candidate with the highest score, $\arg\max_{a\in\mathcal{A}} q(a, D)$ while protecting the privacy of the data subjects whose data is part of D. Classical work in private selection assumes that the score function $q(a, \cdot)$ of all selection candidates a is equally sensitive to changing the data of a single individual. However, in many settings this assumption is false. For example, when performing model selection, more robust models may have lower sensitivity. Intuitively, utilising the fact that some candidates have better than worst-case sensitivity should allow for more accurate differentially private selection mechanisms. Define the *candidate-wise sensitivity* of a score function for candidate $a \in \mathcal{A}$ to be

$$\Delta_a = \max_{D_1, D_2} |q(a, D_1) - q(a, D_2)| \tag{1}$$

where the maximum is over all pairs (D_1, D_2) of datasets that differ on the data of a single individual. In this work we aim to answer the following question: when do algorithms that adapt to heterogeneity in the set $\{\Delta_a\}_{a\in\mathcal{A}}$ outperform standard DP selection algorithms?

When designing differentially private algorithms, it is natural to think that reducing the amount of noise added *anywhere* in an algorithm will improve utility. A surprising finding of our study is that this is not true. There exist settings where adding more noise than necessary actually *improves* performance. To see this, let us consider perhaps the most popular private selection algorithm; Report

[*] Work done while author was intern at Apple

Noisy Max (RNM) Dwork et al. (2006). Define the overall *sensitivity* of the score function to be

$$\Delta = \max_{a \in \mathcal{A}} \Delta_a. \tag{2}$$

Then RNM is defined by

$$M(D) = \operatorname*{argmax}_{a \in \mathcal{A}} \{q(a, D) + z_a\}, \text{ where } z_a \sim \text{Exp}(\epsilon/2\Delta), \tag{3}$$

where ϵ is the privacy parameter and Exp$(\epsilon/2\Delta)$ is the exponential distribution with mean $2\Delta/\epsilon$. RNM adds the same amount of noise to the score of every candidate. Consider an alternative algorithm, RNMH, where the amount of noise added is proportional to the candidate-wise sensitivity of the candidate

$$M(D) = \arg\max_{a \in \mathcal{A}} \{q(a, D) + z_a\}, \text{ where } z_a \sim \text{Exp}(\epsilon/2\Delta_a). \tag{4}$$

Since RNMH adds strictly less noise to the scores of some candidates, we might assume that RNMH always outperforms RNM. Surprisingly, this is not the case. To see this, let us analyse the behaviour of these two algorithms in a specific example. Suppose we have k candidates, $\mathcal{A} = \{a_1, \cdots, a_k\}$, score function q and database D where $\Delta_{a_i} = \Delta_1$ and $q(a_i, D) = q_1$ if $i \leq k/2$, and $\Delta_{a_i} = \Delta_2$ and $q(a_i, D) = q_2$ otherwise. Assume $q_1 < q_2$. RNMH outputs one of the higher scoring candidates whenever

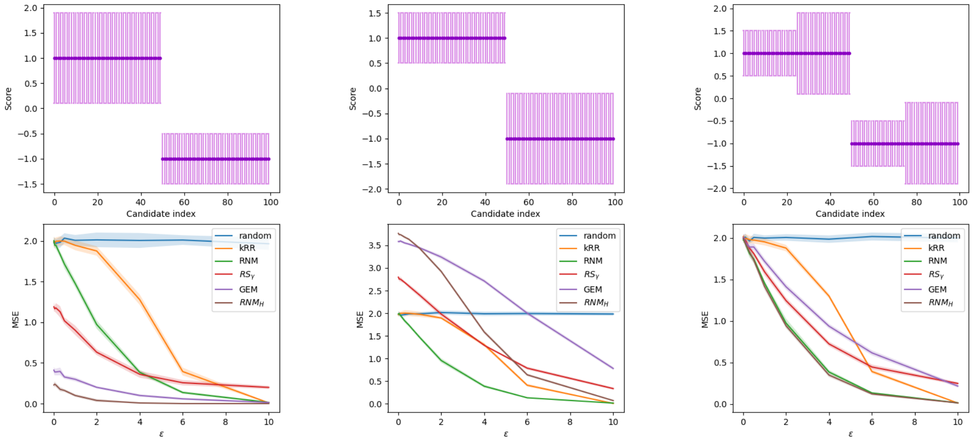
$$\max_{i=1,\cdots,k/2} z_{a_i} - \max_{i=k/2+1,\cdots,k} z_{a_i} \leq q_2 - q_1, \tag{5}$$

where $z_{a_i} \sim$ Exp$(\epsilon/2\Delta_1)$ if $i = 1, \cdots, k/2$ and $z_{a_i} \sim$ Exp$(\epsilon/2\Delta_2)$ otherwise. If $\Delta_1 > \Delta_2$ then $\max_{i=1,\cdots,k/2} z_{a_i}$ stochastically dominates $\max_{i=k/2+1,\cdots,k} z_{a_i}$ (that is, it is more likely to output higher values), so the LHS of eqn (5) is likely to be positive in this setting. In fact, the expectation of the LHS is $H_{k/2}(\Delta_1 - \Delta_2)$ (where $H_{k/2}$ is the $k/2$th harmonic number), which is positive when $\Delta_1 > \Delta_2$ and grows linearly with $\Delta_1 - \Delta_2$. If either k or $\Delta_1 - \Delta_2$ is large, then the probability of RNMH outputting a higher scoring candidate may be less than 50%, which is never the case for RNM. Conversely, when $\Delta_1 < \Delta_2$, we have the opposite effect and RNMH outperforms RNM. We find that the intuition from this example generalises well to more complex distributions of scores and sensitivities; when the scores $q(a, D)$ and sensitivities Δ_a are positively correlated, RNMH generally outperforms RNM. However, when the correlation is negative, RNMH generally performs worse than RNM, and can even perform worse than the algorithm that selects a candidate uniformly at random.

RNMH is the natural extension of RNM to include heterogeneous sensitivities, but unfortunately it is not differentially private (see Section 2 for details). While most of the work on private selection has focused on the homogeneous sensitivities setting (where all the Δ_a sensitivities are assumed to be the same), there have been two main proposals for differentially private selection algorithms that are able to utilise heterogeneity; the *generalised exponential mechanism* (which we'll denote GEM) Raskhodnikova & Smith (2015) and RNM with random stopping (which we'll denote RS$_\gamma$) (an instantiation of Liu & Talwar (2019)). While the relative performance of RNMH and these two algorithms is not straightforward, we expect the intuition for when RNMH outperforms RNM to extend to both of these algorithms. Our empirical experiments, across the settings we tested, consistently support this intuition. In Figure 1 we examine the relative performance of these algorithms in a setting similar to that described in the previous paragraph. We can see that indeed the expected pattern continues. We'll discuss additional results further in Section 2. While prior literature Raskhodnikova & Smith (2015) has given upper bounds on the error in terms of the sensitivity of the optimal candidate, we go beyond this and explore how the distribution of the score-sensitivity pairs $\{(a, \Delta_a)\}$ affects which algorithm is the optimal choice.

Our contributions are as follows:

- We theoretically establish a rule for when RNMH outperforms RNM in the simple setting of bimodal scores and sensitivities setting. Through empirical studies we show that this rule also holds for the ϵ-DP selection algorithms RS$_\gamma$ and GEM. In fact, we see that in the two candidate setting, this effect is exaggerated for both RS$_\gamma$ and GEM.
- We analyze the behaviour of the private selection mechanisms in a range of different settings with a particular focus on the comparison between algorithms that do and do not adapt to heterogeneity in candidate-wise sensitivities. We find that the intuition from the two candidate setting extends to more complex settings with a larger number of candidates; a



(a) Scenario 1: high scores = 1, sensitivities = 1.8; low scores = -1, sensitivities = 1.

(b) Scenario 2: same scores, but the higher left score group now has small sensitivities.

(c) Scenario 3: same scores, but each group has an equal share of large and small sensitivities.

Figure 1: Comparison between algorithms that do and do not adapt to heterogeneity in the candidate-wise sensitivities, in three simple scenarios. Scores and sensitivities here are constant and are shown in the top row, as dark purple dots and light purple vertical lines, respectively. The second row shows mechanism performance (as the mean squared error) as a function of the privacy parameter ϵ.

positive correlation between the scores and the candidate-wise sensitivities indicates that algorithms that utilise heterogeneity will perform better than algorithms that do not. A negative correlation implies the opposite.

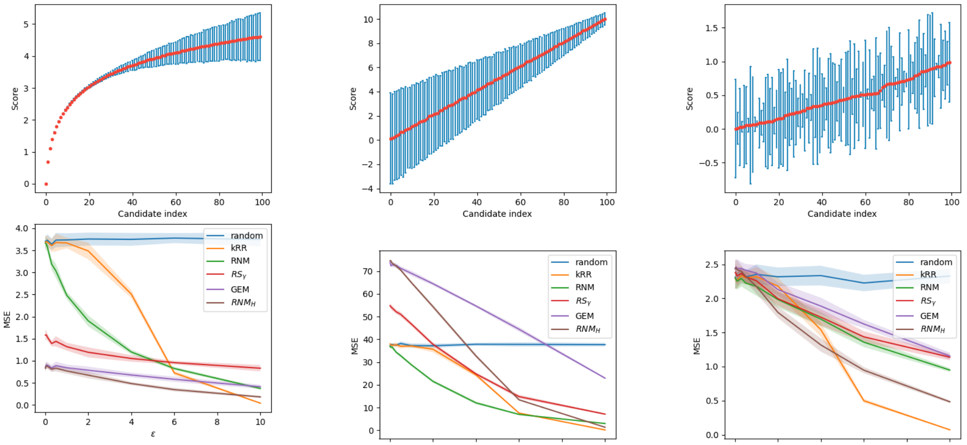
2 Private Selection with Heterogeneous Sensitivities

A Naive Extension of RNM with Heterogeneous Sensitivities is Not Differentially Private. It is tempting to incorporate the heterogeneous sensitivities by running RNM as is, but with noise scaled by the candidate-wise sensitivities resulting in RNMH as described in eqn (4). Unfortunately, this algorithm is not ϵ-differentially-private in general. In fact, there exists pairs of sensitivities Δ_1 and Δ_2 such that this algorithm is not ϵ'-DP for any $\epsilon' > 0$ even when selecting between just two candidates. Consider an example where candidate 1 has score 0 in all databases and so sensitivity $\Delta_1 = 0$. Then let candidate 2 have sensitivity 1 and a score $q_2 = 1/2$ in D_1 and a score $q_2' = -1/2$ in D_2, where D_1 and D_2 are adjacent databases. Then the probability of outputting candidate 1 under dataset D_1 is 0 (since exponential noise is always positive) but under D_2 it is $1 - e^{-\epsilon/4}$, which implies this algorithm is not ϵ'-DP for any ϵ'.

If we use Laplace noise rather than exponential noise in eqn 4 (so $z_a \sim \text{Lap}(\epsilon/\Delta_a)$) then it is easy to see that the resulting algorithm is $k\epsilon$-DP where k is the number of candidates. However, there exists a sequence of sensitivities $\Delta_1, \cdots, \Delta_k$ such that RNM with Laplace noise and heterogeneous sensitivities is not ϵ'-DP for any $\epsilon' < (k-1)\epsilon$. A proof of this appears in the appendix.

Experimental Analysis of When Incorporating Heterogeneity Helps. The experiment results in Figure 2 expand on the intuition we gained in the introduction. In particular, they support the intuition that a positive correlation between the scores and sensitivities indicates that the algorithms RNMH, RS$_\gamma$ and GEM outperform RNM. In the case of negative correlation we find that not only can RNMH, RS$_\gamma$ and GEM perform worse than RNM, they can actually perform worse than the algorithm that selects a candidate uniformly at random.

We create several different scenarios representing different distributions of the set $\{(q(a, D), \Delta_a)\}$. Of course, these scenarios do not provide an exhaustive list of possible distributions of the set $\{(q(a, D), \Delta_a)\}$, but they do provide intuition for realistic scenarios that may arise in practice. For readability, we simplify the notation and replace $q(a, D)$ with q_a throughout this section.



(a) Scenario 4: $\mathcal{N}(\log(a); sorted(\Delta_a^2))$, $\Delta_a \sim \mathcal{N}_{[0.01,0.7]}(0.5; 1)$

(b) Scenario 5: $\mathcal{N}(0.1a; 2.3 - 0.02a)$

(c) Scenario 6: $\mathcal{N}(\mu_a, \Delta_a)$, where $\mu_a \sim U_{[0,1]}$, Δ_a as in (a)

Figure 2: Comparison between algorithms that do and do not adapt to heterogeneity in the candidate-wise sensitivities for three different score and sensitivity distributions with randomness in the scores between trials. *Mean* scores and sensitivities are shown in the top row, as red dots and blue vertical lines, respectively. The second row shows mechanism performance (as the mean squared error) for varying levels of the privacy parameter ϵ.

Let $P_{a,\Delta}$ denote a distribution of candidate scores over experiments from which each score dataset is sampled. We include randomness in the scores to mimic the fact that, in the real world, a domain expert may have an expectation of the relationship between q_a and Δ_a but any particular instantiation will have some variability. Given these score distributions, we generate N sets of the form $\{(q_a, \Delta_a)\}$ as follows. For each candidate a, we generate N samples from $P_{a,\Delta}$, $q_a^1, \cdots, q_a^N$. We then compute the 5th and 95th percentiles, $p_a^{0.05}$ and $p_a^{0.95}$, of the set $\{q_a^1, \cdots, q_a^N\}$ and set $\Delta_a = |p_a^{0.95} - p_a^{0.05}|$. Finally, we truncate each score to be within the range $[p_a^{0.05}, p_a^{0.95}]$. The final N sets are then $\{({q'}_a^i, \Delta_a)\}_{a \in \mathcal{A}}$ for $i \in [N]$, where ${q'}_a^i = \max(p_a^{0.05}, \min(q_a^i, p_a^{0.95}))$. The particulars of each distribution are shown in Figure 2. The following parameters are used, unless specified otherwise: for RS$_\gamma$, $\gamma = 0.05$; for GEM, $\beta = 0.05$. In all scenarios we set $N = 10000$ and $|\mathcal{A}| = 100$. We then run the algorithms on each of these N sets and report the mean squared error (MSE) across experiment runs.

Scenario 4, in Figure 2a, shows an exaggerated view of heterogeneity being useful. Here kRR improves the slowest with ϵ, followed by RNM. Both are outperformed by RNMH, GEM, and RS$_\gamma$ for reasonable ϵ's. Note that, when ϵ is very large, RS$_\gamma$ converges with more error than the other algorithms due to the random stopping, which means it may not see all the candidates. Scenarios 5 and 6 demonstrate that homogeneous noise is preferable when there is no or negative correlation. This is especially true for small ϵ's.

Sensitivities in the wild. Having discussed synthetic scenarios thus far, we use the Yahoo clicks dataset (https://webscope.sandbox.yahoo.com) collected from real user interactions to underscore the practical relevance of our findings. Training a Logistic regression model predicting click through rates of articles for a given user, we obtain the score distributions and truncated candidate-wise sensitivities (Figure 3). Indeed, a positive correlation between the median scores and the sensitivities with correlation coefficient of ≈ 0.71 is observed. Using the intuition we gained from the previous section, we expect the algorithms that utilise heterogeneity in the candidate-wise sensitivities to outperform RNM when privately selecting one of the articles.

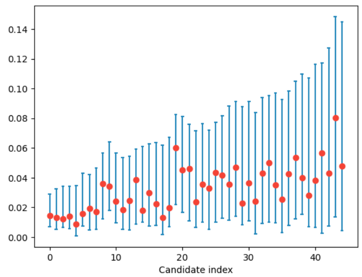


Figure 3: Medians of predicted scores and sensitivities on the Yahoo dataset display a positive correlation.

References

Dwork, C., McSherry, F., Nissim, K., and Smith, A. Calibrating noise to sensitivity in private data analysis. In Halevi, S. and Rabin, T. (eds.), *Theory of Cryptography*, pp. 265–284, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg. ISBN 978-3-540-32732-5.

Liu, J. and Talwar, K. Private selection from private candidates. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, pp. 298–309, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450367059. doi: 10.1145/3313276.3316377. URL `https://doi.org/10.1145/3313276.3316377`.

Raskhodnikova, S. and Smith, A. Efficient lipschitz extensions for high-dimensional graph statistics and node private degree distributions, 2015.

A Counterexample for RNM with Heterogeneous Noise

We showed in Section 2 that RNMH is not differentially private.

In this section, we provide a counterexample to show RNM with Laplacian noise and heterogeneous noise (but otherwise with the original parameters) is not ϵ'-differentially-private for any $\epsilon' < (k-1)\epsilon$, where k is the number of candidates. Consider RNMH with heterogeneous Laplace noise, i.e. the noised scores are $\tilde{q}_a = q_i + z_i$ with $z_i \sim \text{Lap}(\Delta_i/\epsilon)$ and $\epsilon > 0$. Now in dataset D_1 we have $q_i = 0$ for $i = 1, ..., k$ and in dataset D_2 we have $q_1 = 0$ but $q_i = 1$ for $i = 2, ..., k$. Suppose candidate one has sensitivity $\Delta_1 = 0$ and candidates $2, ..., k$ have candidate-wise sensitivities $\Delta_i = 1$.

We now compute the probability of RNMH with Laplacian noise outputting candidate 1 on both datasets. For dataset D_1 we have that

$$\begin{aligned} Pr(M(D_1) = 1) &= Pr(z_i < 0 \text{ for } i = 2, ..., k) && (6) \\ &= \prod_{i=2}^{k} Pr(z_i < 0) && (7) \\ &= \left(\frac{1}{2}\right)^{k-1} && (8) \end{aligned}$$

On the other hand we have that

$$\begin{aligned} Pr(M(D_2) = 1) &= Pr(z_i < -1 \text{ for } i = 2, ..., k) && (9) \\ &= \prod_{i=2}^{k} Pr(z_i < -1) && (10) \\ &= \left(\frac{1}{2}\exp(-\epsilon)\right)^{k-1} && (11) \end{aligned}$$

Thus,

$$\begin{aligned} \frac{Pr(M(D_1) = 1)}{Pr(M(D_2) = 1)} &= \left(\frac{\frac{1}{2}}{\frac{1}{2}\exp(-\epsilon)}\right)^{k-1} && (12) \\ &= \exp((k-1)\epsilon), && (13) \end{aligned}$$

which concludes the counterexample.