Supplements: Differentially Private Selection using Smooth Sensitivity

This provides the omitted proofs in the main paper.

I. Proofs

A. Theorem 3

Theorem 3. *If Z_r are derived from g, the smooth private selection mechanism satisfies $\left(\left(k+\frac{|\mathcal{R}|-1}{2}\cdot l\right)\cdot\epsilon\right)$-differential privacy.*

Proof. We let $\mathcal{R}=\{1,2,\ldots,m\}$ as in the proof of Theorem 2. It is sufficient to show that, for all neighboring datasets $x,y\in D^n$,

$$\begin{aligned}&\Pr[M(x)=1]\\ \leq\ &\exp\left(\left(k+\frac{m-1}{2}\cdot l\right)\cdot\epsilon\right)\cdot\Pr[M(y)=1].\end{aligned}\tag{1}$$

Here, the relations

$$\begin{aligned}\Pr[M(x)=1] = \int_{v\in(-\infty,\infty)} &\Pr\left[Z_1=\frac{\alpha'(v-u(x,1))}{S(x)}\right]\\ &\cdot\Pr\left[Z_2\leq\frac{\alpha'(v-u(x,2))}{S(x)}\right]\\ &\cdots\Pr\left[Z_m\leq\frac{\alpha'(v-u(x,m))}{S(x)}\right]\end{aligned}\tag{2}$$

and

$$\begin{aligned}&\Pr[M(y)=1]\\ \geq\ &\int_{v\in(-\infty,\infty)}\Pr\left[Z_1=\frac{\alpha'(v-u(y,1)+S(x))}{S(y)}\right]\\ &\cdot\Pr\left[Z_2\leq\frac{\alpha'(v-u(x,2))}{S(y)}\right]\\ &\cdots\Pr\left[Z_m\leq\frac{\alpha'(v-u(x,m))}{S(y)}\right]\end{aligned}\tag{3}$$

hold as in the proof of Theorem 2, and in particular, we should note $u(x,1)-u(y,1)+S(x)\geq 0$.

(I) When $S(x)\geq S(y)$:

Because $\forall r\in\{2,3,\ldots,m\}$:

$$\Pr\left[Z_r\leq\frac{\alpha'(v-u(x,r))}{S(y)}\right]\geq\Pr\left[Z_r\leq\frac{\alpha'(v-u(x,r))}{S(x)}\right],\quad[\because\ \text{the property of } g]$$

$$\begin{aligned}(3)\ \geq\ \int_{v\in(-\infty,\infty)}&\Pr\left[Z_1=\frac{\alpha'(v-u(y,1)+S(x))}{S(y)}\right]\\ &\cdot\Pr\left[Z_2\leq\frac{\alpha'(v-u(x,2))}{S(x)}\right]\\ &\cdots\Pr\left[Z_m\leq\frac{\alpha'(v-u(x,m))}{S(x)}\right].\end{aligned}\tag{4}$$

From (2) and (4), the following relation holds:

$$\Pr[M(x)=1]\leq\exp\left(\left(k+\frac{l}{2}\right)\cdot\epsilon\right)\cdot\Pr[M(y)=1].$$

(II) When $S(x)<S(y)$:

Because

$$\begin{aligned}&\Pr\left[Z_1=\frac{\alpha'(v-u(y,1)+S(x))}{S(y)}\right]\\ \geq\ &\Pr\left[Z_1=\frac{\alpha'(v-u(y,1)+S(x))}{S(x)}\right],\\ &[\because\ \text{the property of } g]\end{aligned}$$

$$\begin{aligned}(3)\ \geq\ \int_{v\in(-\infty,\infty)}&\Pr\left[Z_1=\frac{\alpha'(v-u(y,1)+S(x))}{S(x)}\right]\\ &\cdot\Pr\left[Z_2\leq\frac{\alpha'(v-u(x,2))}{S(y)}\right]\\ &\cdots\Pr\left[Z_m\leq\frac{\alpha'(v-u(x,m))}{S(y)}\right].\end{aligned}\tag{5}$$

From (2) and (5), the relation (1) holds. □

B. Theorem 5

Theorem 5. *Given a threshold T, we let the set of x satisfying $LS_f(x)>T$ be U. For any β (>0) satisfying*

$$\beta\leq\min_{x\notin U}\frac{1}{ud(x)}\cdot\ln\left(\frac{GS_f}{LS_f(x)}\right),\tag{6}$$

where $ud(x):=\min_{y\in U}d(y,x)$ represents the shortest distance between x and U, the following function S is a β-smooth upper bound:

$$S(x)=GS_f\cdot e^{-\beta\cdot ud(x)}\ .$$

Proof. From (6), for all $x\notin U$,

$$\begin{aligned}&\beta\leq\frac{1}{ud(x)}\cdot\ln\left(\frac{GS_f}{LS_f(x)}\right)\\ \iff\ &LS_f(x)\leq GS_f\cdot e^{-\beta\cdot ud(x)}.\end{aligned}$$

Therefore, from Definition 4 (for *β-smooth upper bound*), it is sufficient to show that

$$\forall x, y,\ d(x, y) = 1: \quad S(x) \le e^{\beta} \cdot S(y). \tag{7}$$

(I) When $x, y \in U$:
Because $S(x) = S(y) = GS_f$, the relation (7) holds.

(II) When $x \in U$ and $y \notin U$:
Because $S(x) = GS_f$ and

$$S(y) = GS_f \cdot e^{-\beta}, \quad [\because ud(y) = d(x, y) = 1]$$

the relation (7) holds.

(III) When $x \notin U$ and $y \in U$:
Similar to Case (II), $S(x) = GS_f \cdot e^{-\beta}$ and $S(y) = GS_f$; therefore, the relation (7) holds.

(IV) When $x, y \notin U$:
Where X satisfies $X \in U$ and $d(X, x) = ud(x)$, and Y satisfies $Y \in U$ and $d(Y, y) = ud(y)$,

$$\begin{aligned} e^{\beta} \cdot S(y) &= e^{\beta} \cdot GS_f \cdot e^{-\beta \cdot d(Y,y)} \\ &\ge e^{\beta} \cdot GS_f \cdot e^{-\beta \cdot d(X,y)} \quad [\because d(Y, y) \le d(X, y)] \\ &\ge e^{\beta} \cdot GS_f \cdot e^{-\beta (d(X,x) + d(x,y))} \\ &\qquad [\because d(X, y) \le d(X, x) + d(x, y)] \\ &= GS_f \cdot e^{-\beta \cdot d(X,x)} = S(x). \quad [\because d(x, y) = 1] \end{aligned}$$

Therefore, the relation (7) holds. □

C. Lemma 2

Lemma 2. *(b, c) satisfies $LS_{\chi^2_{TDT}}((b, c)) > 6$ when*

$$\begin{aligned} & 0 \le c < \frac{b-8}{7} \ \vee\ 2 \le b < \frac{c+8}{7} \\ \vee\ & 0 \le b < \frac{c-8}{7} \ \vee\ 2 \le c < \frac{b+8}{7}\ . \end{aligned}$$

Proof. $LS_{\chi^2_{TDT}}(x) > 6$ can be satisfied only if $x = (b, c)$ and $y = (b-2, c+2) \vee (b+2, c-2)$.

When $y = (b-2, c+2)$ is possible, $b \ge 2$ and $c \ge 0$. In this case,

$$\begin{aligned} & \chi^2_{TDT}(b, c) - \chi^2_{TDT}(b-2, c+2) \\ = \ & \frac{(b-c)^2}{b+c} - \frac{(b-c-4)^2}{b+c} = \frac{8(b-c-2)}{b+c}. \end{aligned} \tag{8}$$

$|(8)| > 6 \iff c < \frac{b-8}{7} \vee b < \frac{c+8}{7}$; therefore, (b, c) satisfies $LS_{\chi^2_{TDT}}((b, c)) > 6$ when $0 \le c < \frac{b-8}{7} \vee 2 \le b < \frac{c+8}{7}$.

Similarly when $y = (b+2, c-2)$ is possible, we can obtain $0 \le b < \frac{c-8}{7} \vee 2 \le c < \frac{b+8}{7}$. □

D. Lemma 3

Lemma 3. *The Hamming distance for TDT datasets can be*

$$\begin{aligned} & d(T(b, c),\, T(b', c')) \\ = \ & \begin{cases} \left\lceil \frac{|(b+c)-(b'+c')|}{2} \right\rceil & ((b-b') \cdot (c-c') \ge 0) \\ \left\lceil \frac{\max\{|b-b'|,\, |c-c'|\}}{2} \right\rceil & ((b-b') \cdot (c-c') < 0) \end{cases}, \end{aligned}$$

where $T(b, c)$ represents a table that can be formed as the following table:

		Non-Transmitted Allele		Total
		A_1	A_2	
Transmitted Allele	A_1	a	b	$a+b$
	A_2	c	d	$c+d$
Total		$a+c$	$b+d$	$2N$

Proof. The possible changes in (b, c) between neighboring datasets are

$$\begin{aligned} & (b, c) \\ \to\ & (b-2, c),\ (b-2, c+1),\ (b-2, c+2), \\ & (b-1, c-1),\ (b-1, c),\ (b-1, c+1),\ (b-1, c+2), \\ & (b, c-2),\ (b, c-1),\ (b, c),\ (b, c+1),\ (b, c+2), \\ & (b+1, c-2),\ (b+1, c-1),\ (b+1, c),\ (b+1, c+1), \\ & (b+2, c-2),\ (b+2, c-1),\ (b+2, c). \end{aligned}$$

Therefore, the maximum change in the value of $b + c$ is 2; when $(b - b') \cdot (c - c') \ge 0$, the Hamming distance can be

$$d(T(b, c),\, T(b', c')) = \left\lceil \frac{|(b+c) - (b'+c')|}{2} \right\rceil .$$

When $(b - b') \cdot (c - c') < 0$, using the changes $(b, c) \to (b-2, c)$, $(b-2, c+1)$, $(b-2, c+2)$, $(b+2, c-2)$, $(b+2, c-1)$, $(b+2, c)$, the Hamming distance can be

$$d(T(b, c),\, T(b', c')) = \left\lceil \frac{\max\{|b-b'|, |c-c'|\}}{2} \right\rceil .$$

□