Private Causal Inference (Appendix)

Matt J. Kusner
Washington University in St. Louis
mkusner@wustl.edu

Yu Sun
Cornell University
ys646@cornell.edu

Karthik Sridharan
Cornell University
sridharan@cs.cornell.edu

Kilian Q. Weinberger
Cornell University
kqw4@cornell.edu

In this appendix we first reproduce the Propose, Test, Release algorithm of [1] which we use to privatize the Interquartile Range dependence score (IQR), described in Section 4.3 of the main paper. We then prove various claims in the paper.

Algorithm 1 IQR Propose-Test-Release [1]

1: **Input:** data $\mathbf{X} = \{x_1, \ldots, x_m\}$, privacy $\epsilon, \delta > 0$
2: $k = \lfloor \log \mathrm{IQR}(\mathbf{X}) \rfloor$
3: $B_1 = [e^k, e^{k+1})$
4: $B_2 = [e^{k-0.5}, e^{k+0.5})$
5: **for** j = 1,2 **do**
6: $A_j :=$ number of data-points to modify to move $\mathrm{IQR}(\mathbf{X})$ out of interval B_j
7: $R_j = A_j + z$, where $z \sim \mathrm{Lap}(0, \frac{1}{\epsilon})$
8: **if** $R_j > 1 + \log(1/\delta)$ **then**
9: **return** $\log \mathrm{IQR}(\mathbf{X}) + z$, where $z \sim \mathrm{Lap}(0, \frac{1}{\epsilon})$
10: **end if**
11: **end for**
12: **return** $\perp$

Proof of Theorem 2. Let $x_1 \sim \mathrm{Lap}(\mu_1, \sigma)$ and $x_2 \sim \mathrm{Lap}(\mu_2, \sigma)$ be two independent Laplace random variables with $\mu_1 < \mu_2$, then the probability of failure is $\Pr(x_1 > x_2)$. We would like to compute the probability of failure in closed form. We know that by independence, the joint probability is equal to the product of marginal probabilites. We also know that the Laplace cdf. is

$$F(x; \mu, \sigma) = \begin{cases} F_1(x; \mu, \sigma) = \frac{1}{2}\exp(\frac{x-\mu}{\sigma}) & \text{if } x \leq \mu \\ F_2(x; \mu, \sigma) = 1 - \frac{1}{2}\exp(-\frac{x-\mu}{\sigma}) & \text{if } x > \mu \end{cases}$$

where F_1 and F_2 are only defined on the specified domains.

There are six mutually exclusive and collective exhaustive ways for which a failure could happen:

$$(1)\, \mu_1 < x_2 < x_1 < \mu_2$$

$$(2)\, x_2 < \mu_1 < \mu_2 < x_1$$

$$(3)\, \mu_1 < x_2 < \mu_2 < x_1$$

$$(4)\, x_2 < \mu_1 < x_1 < \mu_2$$

$$(5)\, \mu_1 < \mu_2 < x_2 < x_1$$

$$(6)\, x_2 < x_1 < \mu_1 < \mu_2$$

By symmetry of the Laplace distribtuion, we know that $\Pr((3)) = \Pr((4))$ and $\Pr((5)) = \Pr((6))$. Thus we only need to calculate $\Pr((1)), \Pr((2)), \Pr((3))$, and $\Pr((5))$.

$$\begin{aligned}\Pr((1)) &= \int_{\mu_1}^{\mu_2}\int_{x_2}^{\mu_2} p(x_1)p(x_2)dx_1dx_2 \\ &= \int_{\mu_1}^{\mu_2} [F_2(\mu_2; \mu_1, \sigma) - F_2(x_2; \mu_1, \sigma)]p(x_2)dx_2\end{aligned}$$

Now consider the quantity being integrated, which is equal to

$$-\frac{1}{2}\exp(-\frac{\mu_2-\mu_1}{\sigma})p(x_2) + \underbrace{\frac{1}{2}\exp(-\frac{x_2-\mu_1}{\sigma})p(x_2)}_{\star}$$

The right-hand term is,

$$\begin{aligned}\star &= \frac{1}{2}\exp(-\frac{x_2-\mu_1}{\sigma})\frac{1}{2\sigma}\exp(-\frac{\mu_2-x_2}{\sigma}) \quad \text{since } x_2 < \mu_2 \\ &= \frac{1}{4\sigma}\exp(-\frac{\mu_2-\mu_1}{\sigma})\end{aligned}$$

So,

$$\begin{aligned}\Pr((1)) &= \frac{\mu_2-\mu_1}{4\sigma}\exp(-\frac{\mu_2-\mu_1}{\sigma}) \\ &\quad - \frac{1}{2}\exp(-\frac{\mu_2-\mu_1}{\sigma})\int_{\mu_1}^{\mu_2} p(x_2) \\ &= \frac{\mu_2-\mu_1}{4\sigma}\exp(-\frac{\mu_2-\mu_1}{\sigma}) \\ &\quad - \frac{1}{2}\exp(-\frac{\mu_2-\mu_1}{\sigma})[\frac{1}{2} - F_1(\mu_1; \mu_2, \sigma)]\end{aligned}$$

Next, we have that

$$\begin{aligned}\Pr((2)) &= \Pr(x_1 > \mu_2)\Pr(x_2 < \mu_1) \\ &= (1 - F_2(\mu_2; \mu_1, \sigma))F_1(\mu_1; \mu_2, \sigma) \\ &= \frac{1}{2}\exp(-\frac{\mu_2-\mu_1}{\sigma})F_1(\mu_1; \mu_2, \sigma)\end{aligned}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{z_1+z_2-z_3}^{\infty} p(z_1 \mid a,s)p(z_2 \mid b,s)p(z_3 \mid c,s)p(z_4 \mid d,s)dz_4dz_3dz_2dz_1 \tag{1}$$

And similarly,

$$\begin{aligned}\Pr(③) &= \Pr(x_1 > \mu_2)\Pr(\mu_1 < x_2 < \mu_2)\\ &= (1-F_2(\mu_2;\mu_1,\sigma))[\frac{1}{2}-F_1(\mu_1;\mu_2,\sigma)]\\ &= \frac{1}{2}\exp(-\frac{\mu_2-\mu_1}{\sigma})[\frac{1}{2}-F_1(\mu_1;\mu_2,\sigma)]\end{aligned}$$

and as stated, $\Pr(④)$ is the same. Moving on,

$$\begin{aligned}\Pr(⑤) &= \int_{\mu_2}^{\infty}\int_{x_2}^{\infty} p(x_1)p(x_2)dx_1dx_2\\ &= \int_{\mu_2}^{\infty}[1-F_2(x_2;\mu_1,\sigma)]p(x_2)dx_2\\ &= \int_{\mu_2}^{\infty}\frac{1}{2}\exp(-\frac{x_2-\mu_1}{\sigma})\frac{1}{2\sigma}\exp(-\frac{x_2-\mu_2}{\sigma})dx_2\\ &= \frac{1}{4}\int_{\mu_2}^{\infty}\frac{1}{2(\sigma/2)}\exp(-\frac{x_2-(\mu_1+\mu_2)/2}{\sigma/2})dx_2\\ &= \frac{1}{4}(1-F_2(\mu_2;\mu',\sigma'))\\ &= \frac{1}{8}\exp(-\frac{\mu_2-(\mu_1+\mu_2)/2}{\sigma/2}) = \frac{1}{8}\exp(-\frac{\mu_2-\mu_1}{\sigma})\end{aligned}$$

and as stated, $\Pr(⑥)$ is the same. So lastly,

$$\begin{aligned}\Pr(x_1 > x_2) &= 2\Pr(⑤)+2\Pr(③)+\Pr(②)+\Pr(①)\\ =&\frac{1}{4}\exp(-\frac{\mu_2-\mu_1}{\sigma})\\ &+\frac{1}{2}\exp(-\frac{\mu_2-\mu_1}{\sigma})\\ &-\exp(-\frac{\mu_2-\mu_1}{\sigma})F_1(\mu_1;\mu_2,\sigma)\\ &+\frac{1}{2}\exp(-\frac{\mu_2-\mu_1}{\sigma})F_1(\mu_1;\mu_2,\sigma)\\ &+\frac{\mu_2-\mu_1}{4\sigma}\exp(-\frac{\mu_2-\mu_1}{\sigma})\\ &-\frac{1}{4}\exp(-\frac{\mu_2-\mu_1}{\sigma})\\ &+\frac{1}{2}\exp(-\frac{\mu_2-\mu_1}{\sigma})F_1(\mu_1;\mu_2,\sigma)\\ =&\frac{\mu_2-\mu_1+2\sigma}{4\sigma}\exp(-\frac{\mu_2-\mu_1}{\sigma})\end{aligned}$$

This completes the derivation. ∎

Proof of Theorem 4. Given a set of Laplace random variables: $z_1 \sim \mathrm{Lap}(a,s), z_2 \sim \mathrm{Lap}(b,s), z_3 \sim \mathrm{Lap}(c,s), z_4 \sim \mathrm{Lap}(d,s)$ such that $a+b < c+d$, we would like to compute the probability that $z_1+z_2 < z_3+z_4$. Let $p(x \mid \mu,\sigma)$ be the pdf of the Laplace distribution $\mathrm{Lap}(\mu,\sigma)$. Computing the above probability requires evaluating the expression in eq. (1). Similar to the above proof, we can compute this integral by enumerating all of the possible cases, which gives the stated result. ∎

Proof of Theorem 5.

$$\begin{aligned}|r'_{i,Y}-\tilde{r}'_{i,Y}| &= \left|\mathbf{w}^\top\phi(X)-\tilde{\mathbf{w}}^\top\phi(X)\right| &(2)\\ &\le \|\mathbf{w}-\tilde{\mathbf{w}}\|_{\mathcal{H}}\|\phi(X)\|_{\mathcal{H}} \le \|\mathbf{w}-\tilde{\mathbf{w}}\|_{\mathcal{H}} &(3)\end{aligned}$$

In the above we used the fact that $\|\phi(X)\|_{\mathcal{H}} = \sqrt{K(x,x)} \le 1$. On the other hand note that $\mathbf{w}$ is the minimizer of regularized objective on data set $(x_1,y_1),\ldots,(x_n,y_n)$ and $\tilde{\mathbf{w}}$ is the minimizer on set $(x_1,y_1),\ldots,(x_{n-1},y_{n-1}),(x'_n,y'_n)$ (we assume the last coordinate is the one that is changes w.l.o.g.). By strong convexity of the regularized objective we have,

$$\begin{aligned}\frac{\lambda}{2}\|\mathbf{w}-\tilde{\mathbf{w}}\|^2 &\le \frac{\lambda}{2}\|\tilde{\mathbf{w}}\|^2_{\mathcal{H}}+\frac{1}{n}\sum_{i=1}^{n}(\tilde{\mathbf{w}}^\top\phi(x_i)-y_i)^2\\ &\quad-\frac{\lambda}{2}\|\mathbf{w}\|^2_{\mathcal{H}}-\frac{1}{n}\sum_{i=1}^{n}(\mathbf{w}^\top\phi(x_i)-y_i)^2\\ &\le \frac{\lambda}{2}\|\tilde{\mathbf{w}}\|^2_{\mathcal{H}}+\frac{1}{n}\sum_{i=1}^{n-1}(\tilde{\mathbf{w}}^\top\phi(x_i)-y_i)^2+(\tilde{\mathbf{w}}^\top\phi(\tilde{x}_n)-\tilde{y}_n)^2\\ &\quad-\frac{\lambda}{2}\|\mathbf{w}\|^2_{\mathcal{H}}-\frac{1}{n}\sum_{i=1}^{n-1}(\mathbf{w}^\top\phi(x_i)-y_i)^2-(\mathbf{w}^\top\phi(\tilde{x}_n)-\tilde{y}_n)^2\\ &\quad+\frac{1}{n}(\tilde{\mathbf{w}}^\top\phi(x_n)-y_n)^2-(\mathbf{w}^\top\phi(x_n)-y_n)^2\\ &\quad-\frac{1}{n}(\tilde{\mathbf{w}}^\top\phi(\tilde{x}_n)-\tilde{y}_n)^2+(\mathbf{w}^\top\phi(\tilde{x}_n)-\tilde{y}_n)^2\\ &\le \frac{2}{n}\sup_{x,y\in[-1,1]}\left((\tilde{\mathbf{w}}^\top\phi(x)-y)^2-(\mathbf{w}^\top\phi(x)-y)^2\right)\\ &\le \frac{2}{n}\|\tilde{\mathbf{w}}-\mathbf{w}\|\times(\|\tilde{\mathbf{w}}\|+\|\mathbf{w}\|+2)\end{aligned}$$

Now note that since $0 \in \mathcal{H}$ we can conclude that,

$$\|\mathbf{w}\| \le \frac{1}{\sqrt{\lambda}}$$

(The above is got by plugging in the 0 in the regularized objective which yields a value of 1 and since loss is non-negative, we can conclude that the norm of the minimizer of the regularized objective is atmost $1/\sqrt{\lambda}$. Plugging this in yields:

$$\|\mathbf{w}-\tilde{\mathbf{w}}\| \le \frac{8}{\lambda^{3/2}n}$$

Plugging this in Eq. 2 yields the theorem. ∎

Proof of tighter HSIC bound. Let D be the original dataset and D' be the dataset with one column modified. We subscript $HSIC$ with l and k implicitly. This is the quantity of interest

$$|H\hat{S}IC(D) - H\hat{S}IC(D')| = \frac{1}{(N-1)^2}|tr(K'HL'H) - tr(KHLH)|$$

Pulling out the constant, we have

$$\begin{aligned}
&(N-1)^2|H\hat{S}IC(D) - H\hat{S}IC(D')| \\
&= |tr(K'HL'H) - tr(KHLH)| \\
&= |tr((K'HL' - KHL)H)| && \text{linearity of trace} \\
&= |tr(H(K'HL' - KHL))| && \text{cyclicity of trace}
\end{aligned}$$

Let $\mathbf{1}$ be the square matrix of $ones(N)$. We know that since $H = I - \frac{1}{N}\mathbf{1}$ by definition,

$$H(K'HL' - KHL) = (K'HL' - KHL) - \frac{1}{N}\mathbf{1}(K'HL' - KHL)$$

so we have that

$$\begin{aligned}
&(N-1)^2|H\hat{S}IC(D) - H\hat{S}IC(D')| \\
&= |tr(K'HL' - KHL) - \frac{1}{N}tr(\mathbf{1}(K'HL' - KHL))|
\end{aligned} \tag{4}$$

Next, we need three identities. Let $sum(A) = \sum_{i,j} A$, then

Identity 1: $tr(\mathbf{1}A) = sum(A)$

Identity 2: $tr(\mathbf{1}A\mathbf{1}B) = sum(A)sum(B)$

Identity 3: $sum(AB) = sum(BA)$

Where Identity 3 holds only for symmetric matrices. Identity 3 is obvious since $AB = (BA)^T$ and $sum(C) = sum(C^T)$, while the first two can be proven by expanding out the matrices and using the row-column rule, or just trying random matrices on MATLAB until you believe that it works. I did both, they are sure to be correct.
And again from the definition of H, we know that

$$KHL = K(L - \frac{1}{N}\mathbf{1}L) = KL - \frac{1}{N}K\mathbf{1}L$$

so

$$K'HL' - KHL = (K'L' - \frac{1}{N}K'\mathbf{1}L') - (KL - \frac{1}{N}K\mathbf{1}L)$$

Now we continue our derivation of eq. (4)

$$(N-1)^2|H\hat{S}IC(D) - H\hat{S}IC(D')| = \underbrace{|tr(K'HL' - KHL)}_{\star} - \underbrace{\frac{1}{N}sum(K'HL' - KHL)|}_{\diamond}$$

We can rewrite each term $\star$ and $\diamond$ using our traces identities as follows,

$$\star = [tr(K'L') - \frac{1}{N}sum(L'K')] - [tr(KL) - \frac{1}{N}sum(LK)]$$

$$\begin{aligned}
\diamond &= \frac{1}{N}[sum(K'L' - \frac{1}{N}K'\mathbf{1}L') - sum(KL - \frac{1}{N}K\mathbf{1}L)] \\
&= \frac{1}{N}[sum(K'L') - \frac{1}{N}sum(K'\mathbf{1}L')] \\
&\quad - \frac{1}{N}[sum(KL) - \frac{1}{N}sum(K\mathbf{1}L)]
\end{aligned}$$

By identity 3, we see that the $sum(KL)$ and $sum(LK)$ as well as the $sum(K'L')$ and $sum(L'K')$ terms in $\star$ and $\diamond$ are identical. Thus we are left with

$$\begin{aligned}
&(N-1)^2|H\hat{S}IC(D) - H\hat{S}IC(D')| = \star - \diamond \\
&= |[tr(K'L') - tr(KL)] \\
&\quad - \frac{2}{N}[sum(K'L') - sum(KL)] \\
&\quad + \frac{1}{N}[\frac{1}{N}sum(K'\mathbf{1}L') - \frac{1}{N}sum(K\mathbf{1}L)]| \\
&= |[sum(K'.*L') - sum(K.*L)] \\
&\quad - \frac{2}{N}[sum(K'L') - sum(KL)] \\
&\quad + \frac{1}{N^2}[sum(K')sum(L') - sum(K)sum(L)]|
\end{aligned} \tag{5}$$

where the last line comes from applying Identity 1 backwards so we have, for example, $tr(\mathbf{1}K\mathbf{1}L)$, then applying Identity 2. We use MATLAB© notation $.*$ for the element-wise product of two matrices.

We bound eq. (5) by the triangle inequality,

$$\begin{aligned}
&(N-1)^2|H\hat{S}IC(D) - H\hat{S}IC(D')| \\
&\le |sum(K'.*L') - sum(K.*L)| && (1) \\
&+ \frac{2}{N}|sum(K'L') - sum(KL)| && (2) \\
&+ \frac{1}{N^2}|sum(K')sum(L') - sum(K)sum(L)| && (3)
\end{aligned}$$

$$\begin{aligned}
&(N-1)^2|H\hat{S}IC(D) - H\hat{S}IC(D')| \\
&\le [(1) + (2) + (3)] \le \max_{K,K',L,L'} (1) + \max_{K,K',L,L'} (2) + \max_{K,K',L,L'} (3)
\end{aligned}$$

Recall that the kernels k and l are bounded by 1. And that the kernel pairs K, K' and L, L' differ in at most one row and column. Thus, for (1), it is clear that the maximum occurs when a row and column c (no matter what c is) is changed from all 0 to 1 in both L and K, so $\max_{K,K',L,L'} (1) = 2N - 1$
For (3), the maximum occurs at exactly same the point as (1), and the value achieved is

$$\frac{1}{N^2}[N^4 - (N^2 - 2N + 1)^2] \le 4N - 5$$

for $N > 3$. For ②, applying the row-column rule and reasoning on small matrices inductively suggest that the maximum is also achieved when we change one row and column of L and K from all 0 to 1, and is thus

$$\frac{2}{N}[N^2 + (N-1)(2N-1)] \leq 6N - 5$$

for $N \geq 2$. As the argmax of all three terms coincide, we have that

$$\max_C [① + ② + ③] = \max_C ① + \max_C ② + \max_C ③$$

Therefore, we have derived that for all practical purposes, the overall bound is $\frac{12N-11}{N^2-1}$. ■

References

[1] Dwork, Cynthia and Lei, Jing. Differential privacy and robust statistics. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pp. 371–380. ACM, 2009.