Online Appendix for “Information Design for Differential Privacy”

Ian M. Schmutte and Nathan Yoder*

June 29, 2024

S.1 Differential Privacy and Learning

Here, we offer an interpretation of differential privacy as a bound on Bayesian updating. Proposition S.1 shows that differential privacy is equivalent to a bound on the amount that learning the mechanism’s output can cause an observer who believes respondents’ types are independent to update their beliefs about a specific respondent’s type θ_n. In particular, differential privacy limits the proportional change in the odds that the respondent is type 1. The argument mirrors that of Theorem 6.1 in Kifer and Machanavajjhala (2014). However, Proposition S.1 differs in that it bounds the change in the odds that the respondent has one type instead of another, rather than the change in the odds that it has a certain type instead of being absent from the data altogether.

Proposition S.1 (Interpretations of Differential Privacy)**.** *The following are equivalent:*

i. *(S, m) is ϵ-differentially private.*

ii. *If an agent’s prior $\hat{\pi}_0 \in \Delta(\Theta)$ is a product distribution which places positive probability on both $\theta_n = 1$ and $\theta_n = 0$, then after observing a realization s from (S, m), the log odds of the event $\{\theta : \theta_n = 1\}$ under the agent’s posterior $\hat{\pi}$ can differ by no more than ϵ from its log odds under $\hat{\pi}_0$:*

$$\left|\log\left(\frac{\hat{\pi}(\{\theta : \theta_n = 1\})}{\hat{\pi}(\{\theta : \theta_n = 0\})}\right) - \log\left(\frac{\hat{\pi}_0(\{\theta : \theta_n = 1\})}{\hat{\pi}_0(\{\theta : \theta_n = 0\})}\right)\right| \leq \epsilon.$$

*Schmutte: University of Georgia, Terry College of Business, Department of Economics; E-mail: schmutte@uga.edu. Yoder: University of Georgia, Terry College of Business, Department of Economics; E-mail: nathan.yoder@uga.edu.

Proof. ((i)⇒(ii)): Suppose that (S, m) is differentially private and that an agent's prior $\hat{\pi}_0$ is a product distribution. Then for $t \in \{0,1\}$ we can write

$$\hat{\pi}(\{\theta : \theta_n = t\}) = \frac{\sum_{\theta:\theta_n=t} m(s|\theta)\hat{\pi}_0(\theta)}{\sum_{\theta\in\{0,1\}^N} m(s|\theta)\hat{\pi}_0(\theta)} = \frac{\sum_{\theta:\theta_n=t} m(s|\theta)\hat{\pi}_0(\{\hat{\theta} : \hat{\theta}_n = t\})\hat{\pi}_0(\{\hat{\theta} : \hat{\theta}_{-n} = \theta_{-n}\})}{\sum_{\theta\in\{0,1\}^N} m(s|\theta)\hat{\pi}_0(\theta)}$$

Hence

$$\left|\log\left(\frac{\hat{\pi}(\{\theta : \theta_n = 1\})}{\hat{\pi}(\{\theta : \theta_n = 0\})}\right) - \log\left(\frac{\hat{\pi}_0(\{\theta : \theta_n = 1\})}{\hat{\pi}_0(\{\theta : \theta_n = 0\})}\right)\right| = \left|\log\left(\frac{\sum_{\theta:\theta_n=1} m(s|\theta)\hat{\pi}_0(\{\hat{\theta} : \hat{\theta}_{-n} = \theta_{-n}\})}{\sum_{\theta:\theta_n=0} m(s|\theta)\hat{\pi}_0(\{\hat{\theta} : \hat{\theta}_{-n} = \theta_{-n}\})}\right)\right|$$
$$= \left|\log\left(\frac{\sum_{\theta_{-n}\in\{0,1\}^{N-1}} m(s|(1,\theta_{-n}))\hat{\pi}_0(\{\hat{\theta} : \hat{\theta}_{-n} = \theta_{-n}\})}{\sum_{\theta_{-n}\in\{0,1\}^{N-1}} m(s|(0,\theta_{-n}))\hat{\pi}_0(\{\hat{\theta} : \hat{\theta}_{-n} = \theta_{-n}\})}\right)\right|.$$

Now we have

$$\left(\min_{\theta_{-n}\in\{0,1\}^{N-1}} \left\{\frac{m(s|(1,\theta_{-n}))}{m(s|(0,\theta_{-n}))}\right\}\right)\left(\sum_{\theta_{-n}\in\{0,1\}^{N-1}} m(s|(0,\theta_{-n}))\hat{\pi}_0(\{\hat{\theta} : \hat{\theta}_{-n} = \theta_{-n}\})\right)$$
$$\leq \sum_{\theta_{-n}\in\{0,1\}^{N-1}} m(s|(1,\theta_{-n}))\hat{\pi}_0(\{\hat{\theta} : \hat{\theta}_{-n} = \theta_{-n}\})$$
$$\leq \left(\max_{\theta_{-n}\in\{0,1\}^{N-1}} \left\{\frac{m(s|(1,\theta_{-n}))}{m(s|(0,\theta_{-n}))}\right\}\right)\left(\sum_{\theta_{-n}\in\{0,1\}^{N-1}} m(s|(0,\theta_{-n}))\hat{\pi}_0(\{\hat{\theta} : \hat{\theta}_{-n} = \theta_{-n}\})\right),$$

and so

$$\min_{\theta_{-n}\in\{0,1\}^{N-1}} \left\{\frac{m(s|(1,\theta_{-n}))}{m(s|(0,\theta_{-n}))}\right\} \leq \frac{\sum_{\theta_{-n}\in\{0,1\}^{N-1}} m(s|(1,\theta_{-n}))\hat{\pi}_0(\{\hat{\theta}:\hat{\theta}_{-n}=\theta_{-n}\})}{\sum_{\theta_{-n}\in\{0,1\}^{N-1}} m(s|(0,\theta_{-n}))\hat{\pi}_0(\{\hat{\theta}:\hat{\theta}_{-n}=\theta_{-n}\})} \leq \max_{\theta_{-n}\in\{0,1\}^{N-1}} \left\{\frac{m(s|(1,\theta_{-n}))}{m(s|(0,\theta_{-n}))}\right\};$$
$$\Rightarrow \log\left(\frac{\sum_{\theta_{-n}\in\{0,1\}^{N-1}} m(s|(1,\theta_{-n}))\hat{\pi}_0(\{\hat{\theta}:\hat{\theta}_{-n}=\theta_{-n}\})}{\sum_{\theta_{-n}\in\{0,1\}^{N-1}} m(s|(0,\theta_{-n}))\hat{\pi}_0(\{\hat{\theta}:\hat{\theta}_{-n}=\theta_{-n}\})}\right)$$
$$\leq \max\left\{\left|\log\left(\min_{\theta_{-n}\in\{0,1\}^{N-1}} \left\{\frac{m(s|(1,\theta_{-n}))}{m(s|(0,\theta_{-n}))}\right\}\right)\right|, \left|\log\left(\max_{\theta_{-n}\in\{0,1\}^{N-1}} \left\{\frac{m(s|(1,\theta_{-n}))}{m(s|(0,\theta_{-n}))}\right\}\right)\right|\right\} \leq \epsilon,$$

as desired.

((ii)⇒(i)): Let $n \in \{1, \ldots, N\}$ and let $\theta, \theta' \in \Theta$ be such that $\theta_{-n} = \theta'_{-n}$. Let $\hat{\pi}_0$ be such that $\hat{\pi}_0(\{\hat{\theta} : \hat{\theta}_{-n} = \theta_{-n}\}) = 1$ and $\hat{\pi}_0(\{\hat{\theta} : \hat{\theta}_n = 1\}) \in (0,1)$. Then for $t \in \{0,1\}$, $\hat{\pi}_0(\{\theta : \theta_n = t\}) = \hat{\pi}_0(\theta_{-n}, t)$ and $\hat{\pi}(\{\theta : \theta_n = t\}) = \frac{m(s|(t,\theta_{-n}))\hat{\pi}_0(\theta_{-n},t)}{m(s|(t,\theta_{-n}))\hat{\pi}_0((t,\theta_{-n}))+m(s|(1-t,\theta_{-n}))\hat{\pi}_0((1-t,\theta_{-n}))}$. Hence

$$\epsilon \geq \left|\log\left(\frac{\hat{\pi}(\{\theta : \theta_n = 1\})}{\hat{\pi}(\{\theta : \theta_n = 0\})}\right) - \log\left(\frac{\hat{\pi}_0(\{\theta : \theta_n = 1\})}{\hat{\pi}_0(\{\theta : \theta_n = 0\})}\right)\right| = \left|\log\left(\frac{m(s|(1,\theta_{-n}))}{m(s|(0,\theta_{-n}))}\right)\right| = \left|\log\left(\frac{m(s|\theta')}{m(s|\theta)}\right)\right|.$$

Since this holds for any $s \in S$, (i) follows. □

S.2 Permutation-Invariant Mechanisms

Here, we prove Proposition 5 and characterize differential privacy for permutation-invariant mechanisms.

In what follows, let $\sim$ be the permutation equivalence relation on Θ: $\theta \sim \theta' \Leftrightarrow \theta$ is a permutation of θ'. By definition of ω_θ, $\theta \sim \theta'$ implies $\omega_\theta = \omega_{\theta'}$. Let $\mathcal{C}$ denote the collection of equivalence classes of $\sim$, and for each $C \in \mathcal{C}$, let ω_C denote the common value of ω_θ across all $\theta \in C$. For each $\theta \in \Theta$, let C_θ denote its $\sim$-equivalence class. We say that two equivalence classes $C, C' \in \mathcal{C}$ are *adjacent* if there exist $\theta \in C$ and $\theta' \in C'$ such that for some i, $\theta_{-i} = \theta'_{-i}$ but $\theta_i \neq \theta'_i$. Note that for any such pair θ, θ', θ_i must take the same value, which we denote $t(C, C')$; let $n(C, C')$ denote the number of entries of θ that take this value.

For a permutation-invariant mechanism (S, m), there is a function $\rho : \mathcal{C} \to \Delta(S)$ such that for every $\theta \in C \in \mathcal{C}$, $m(\cdot|\theta) = \rho(\cdot|C)$; we abuse notation and write (S, ρ) to denote such a mechanism. Define the projection operator $P_\mathcal{C} : \Delta(\Theta) \to \Delta(\mathcal{C})$ by $P_\mathcal{C}\pi(C) = \sum_{\theta \in C} \pi(\theta)$, and the common prior about the permutation class as $\beta_0 \equiv P_\mathcal{C}\pi_0$. Then we can characterize differential privacy for permutation-invariant mechanisms in terms of the posterior beliefs they induce about the permutation equivalence class C, as follows.

Proposition S.2 (Differential Privacy for Permutation-Invariant Mechanisms)**.** *Suppose (S, σ) is an oblivious data publication mechanism. Then the following are equivalent:*

i. (S, ρ) is ϵ-differentially private.

ii. $\left|\log\left(\frac{\rho(s|C)}{\rho(s|C')}\right)\right| \leq \epsilon$ for each adjacent $C, C' \in \mathcal{C}$.

iii. For each posterior belief about the permutation class $\beta \in \Delta(\mathcal{C})$ induced by (S, ρ),

$$\left|\log\left(\frac{\beta(C)}{\beta(C')}\right) - \log\left(\frac{\beta_0(C)}{\beta_0(C')}\right)\right| \leq \epsilon \text{ for each adjacent } C, C' \in \mathcal{C}. \tag{1}$$

Proof. ((i)⇒(ii)) For each adjacent $C, C' \in \mathcal{C}$, by definition there exist $\theta \in C$, $\theta' \in C'$, and $i \in \{1, \ldots, N\}$ such that $\theta_i \neq \theta'_i$ and $\theta_{-i} = \theta'_{-i}$. Then since (S, ρ) is ϵ-differentially private, for each $s \in S$, $|\log(\rho(s|C)/\rho(s|C'))| = |\log(\rho(s|C_\theta)/\rho(s|C_{\theta'}))| \leq \epsilon,$; (ii) follows.

((ii)⇒(i)) If $\theta, \theta' \in \{0, \ldots, T\}^N$ are such that $\theta_{-i} = \theta'_{-i}$ for some i, then either $\theta = \theta'$, in which case (1) holds trivially, or C_θ and $C_{\theta'}$ are adjacent, in which case (ii) implies that for each $s \in S$, $|\log(\rho(s|C_\theta)/\rho(s|C_{\theta'}))| = |\log(\rho(s|C_{\theta'})/\rho(s|C_\theta))| \leq \epsilon$, and hence, since (S, ρ) is permutation-invariant, (1).

((ii)⇔(iii)) Follows from Bayes' rule, since

$$\frac{\beta(C)}{\beta(C')} = \frac{\rho(s|C)\beta_0(C)}{\sum_{X \in \mathcal{C}} \rho(s|X)\beta_0(X)} \Bigg/ \frac{\rho(s|C')\beta_0(C')}{\sum_{X \in \mathcal{C}} \rho(s|X)\beta_0(X)} = \frac{\rho(s|C)}{\rho(s|C')}\frac{\beta_0(C)}{\beta_0(C')}.$$

□

Let $K_{\mathcal{C}}(\epsilon, \beta_0)$ denote the set of posterior beliefs about the permutation equivalence class $\beta \in \Delta(\mathcal{C})$ that satisfy (1).

Lemma S.1. *The following are equivalent:*

i. If the distribution $\xi \in \Delta(\Delta(\mathcal{C}))$ of posterior beliefs about the permutation equivalence class can be induced by an ϵ-differentially private mechanism, it can be induced by an ϵ-differentially private oblivious mechanism.

ii. $K_{\mathcal{C}}(\epsilon, \beta_0) = P_{\mathcal{C}} K(\epsilon, \pi_0)$.

Proof. Follows identically to the proof of Lemma 8, relying on Proposition S.2 instead of Proposition 3. □

Proof of Proposition 5 (Permutation-Invariant Mechanisms) Suppose $\pi \in K(\epsilon, \pi_0)$. Then for each adjacent $C, C' \in \mathcal{C}$, we have

$$\begin{aligned} P_{\mathcal{C}}\mu(C) &= \sum_{\theta \in C} \pi(\theta) = \sum_{\theta \in C} \frac{1}{n(C,C')} \sum_{i:\theta_i = t(C,C')} \pi(\theta) \\ &= \frac{1}{n(C,C')} \sum_{i=1}^{N} \sum_{\substack{\theta:\theta_i = t(C,C'),\\ \theta \in C}} \pi(\theta) = \frac{1}{n(C,C')} \sum_{n=1}^{N} \sum_{\substack{\theta':\theta'_i = t(C',C),\\ \theta' \in C'}} \pi((t(C,C'), \theta'_{-i})) \\ &= \frac{1}{n(C,C')} \sum_{\theta' \in C'} \sum_{i:\theta'_i = t(C',C)} \pi((t(C,C'), \theta'_{-i})). \end{aligned}$$

Since respondents are anonymous, $\pi_0(\theta) = \pi_0(\theta')$ whenever $\theta, \theta' \in C$. It follows that for each $\theta \in \Theta$, $\pi_0(\theta) = \beta_0(C_\theta)/|C_\theta|$. Moreover, note that for each $\theta \in \Theta$, $|C_\theta| = N!/\prod_{k=1}^{T} |\{i | \theta_i = k\}|$. Consequently, for each θ' with $\theta_{-i} = \theta'_{-i}$ for some i, $|C_\theta| = \frac{n(C_{\theta'}, C_\theta)}{n(C_\theta, C_{\theta'})} |C_{\theta'}|$.

Then since $\pi \in K(\epsilon, \pi_0)$, for each $\theta' \in \Theta$, each C that is adjacent to $C_{\theta'}$, each i, and each $s \in S$, we have

$$\begin{aligned} e^{-\epsilon} \pi(\theta) \frac{\pi_0((t(C, C_{\theta'}), \theta_{-n}))}{\pi_0(\theta)} &\leq \pi((t(C, C_{\theta'}), \theta'_{-i})) \leq e^{\epsilon} \pi(\theta') \frac{\pi_0((t(C, C_{\theta'}), \theta'_{-i}))}{\pi_0(\theta')} \\ e^{-\epsilon} \pi(\theta') \frac{\beta_0(C)|C_{\theta'}|}{\beta_0(C_{\theta'})|C|} &\leq \pi((t(C, C_{\theta'}), \theta'_{-i})) \leq e^{\epsilon} \pi(\theta') \frac{\beta_0(C)|C_{\theta'}|}{\beta_0(C_{\theta'})|C|} \\ e^{-\epsilon} \pi(\theta) \frac{\beta_0(C) n(C, C_{\theta'})}{\beta_0(C_{\theta'}) n(C_{\theta'}, C)} &\leq \pi((t(C, C_{\theta'}), \theta'_{-i})) \leq e^{\epsilon} \pi(\theta') \frac{\beta_0(C) n(C, C_{\theta'})}{\beta_0(C_{\theta'}) n(C_{\theta'}, C)} \end{aligned}$$

Hence, for each adjacent $C, C' \in \mathcal{C}$, we have

$$e^{-\epsilon} \sum_{\theta' \in C'} \frac{1}{n(C', C)} \sum_{i:\theta'_i = t(C', C)} \pi(\theta') \frac{\beta_0(C)}{\beta_0(C')} \leq P_{\mathcal{C}}\pi(C) \leq e^{\epsilon} \sum_{\theta' \in C'} \frac{1}{n(C', C)} \sum_{i:\theta'_i = t(C', C)} \pi(\theta') \frac{\beta_0(C)}{\beta_0(C')}$$

$$e^{-\epsilon} \sum_{\theta' \in C'} \pi(\theta') \frac{\beta_0(C)}{\beta_0(C')} \leq P_{\mathcal{C}}\pi(C) \leq e^{\epsilon} \sum_{\theta' \in C'} \pi(\theta') \frac{\beta_0(C)}{\beta_0(C')}$$

$$e^{-\epsilon} P_{\mathcal{C}}\pi(C') \frac{\beta_0(C)}{\beta_0(C')} \leq P_{\mathcal{C}}\pi(C) \leq e^{\epsilon} P_{\mathcal{C}}\pi(C') \frac{\beta_0(C)}{\beta_0(C')},$$

and so $P_{\mathcal{C}}\pi \in K_{\mathcal{C}}(\epsilon, \beta_0)$.

Hence, $P_{\mathcal{C}}K(\epsilon, \pi_0) \subseteq K_{\mathcal{C}}(\epsilon, \beta_0)$. And since permutation-invariant ϵ-differentially private mechanisms are a subset of all ϵ-differentially private mechanisms, by Proposition S.2 and Lemma 4, $K_{\mathcal{C}}(\epsilon, \beta_0) \subseteq P_{\mathcal{C}}K(\epsilon, \pi_0)$. So $P_{\mathcal{C}}K(\epsilon, \pi_0) = K_{\mathcal{C}}(\epsilon, \beta_0)$; the statement follows by Lemma S.1. □

References

KIFER, D. AND A. MACHANAVAJJHALA (2014): "Pufferfish: A Framework for Mathematical Privacy Definitions," *ACM Transactions on Database Systems (TODS)*, 39, 1–36.