Differentially Private Bayesian Inference for Generalized Linear Models: Supplementary Materials

1 Examples of moment calculation using Isserlis' Theorem

Theorem 1.1. *Isserlis' theorem [1]. Let $\boldsymbol{x} \sim \mathcal{N}^d(\mathbf{0}, \boldsymbol{\Sigma})$ be a $d-$dimensional random variable.*

$$\mathbb{E}[x_1, \cdots, x_d] = \sum_{p \in P_d^2} \prod_{\{i,j\} \in p} \mathbb{E}[x_i x_j] = \sum_{p \in P_d^2} \prod_{\{i,j\} \in p} \Sigma_{ij} \tag{1}$$

Where d is assumed to be an even number and P is the set of all possible ways of partitioning $\{1, \cdots, d\}$ in to pairs $\{i, j\}$. For odd d's, $\mathbb{E}[x_1, \cdots, x_d] = 0$.

For example, two 4th order moments are computed using the Isserlis' theorem below.

$$\begin{aligned}
\mathbb{E}[x_1^2 x_2^2] &= \mathbb{E}[x_1 x_1 x_2 x_2] \\
&= \Sigma_{11}\Sigma_{22} + \Sigma_{12}\Sigma_{12} + \Sigma_{12}\Sigma_{12} \\
&= \Sigma_{11}\Sigma_{22} + 2\Sigma_{12}\Sigma_{12} \\
\mathbb{E}[x_1^2 x_2 x_3] &= \mathbb{E}[x_1 x_1 x_2 x_3] \\
&= \Sigma_{11}\Sigma_{23} + \Sigma_{12}\Sigma_{13} + \Sigma_{13}\Sigma_{12} \\
&= \Sigma_{11}\Sigma_{23} + 2\Sigma_{12}\Sigma_{13}
\end{aligned}$$

2 Poisson regression with softplus as the link function

s We use the softplus link function $\mu = \mathbb{E}[y|\boldsymbol{x}, \boldsymbol{\theta}] = \log(1 + \exp(\boldsymbol{x}^T\boldsymbol{\theta}))$. The probability mass function is given by

$$\Pr[y|\boldsymbol{x}, \boldsymbol{\theta}] = \frac{\mu^y \exp(-\mu)}{y!} = \frac{(\log(1 + \exp(\boldsymbol{x}^T\boldsymbol{\theta})))^y \exp(-\log(1 + \exp(\boldsymbol{x}^T\boldsymbol{\theta})))}{y!} \tag{2}$$

And the log-likelihood is given by

$$\log\Big[\Pr[y|\boldsymbol{x}, \boldsymbol{\theta}]\Big] \propto y \log(\log(1 + \exp(\boldsymbol{x}^T\boldsymbol{\theta}))) - \log(1 + \exp(\boldsymbol{x}^T\boldsymbol{\theta})) \tag{3}$$

Now we find the normal approximation to each term in equation 3.

2.1 $\log(1 + \exp(\boldsymbol{x}^T\boldsymbol{\theta}))$

The second part of the likelihood i.e. $\log(1 + \exp(\boldsymbol{x}^T\boldsymbol{\theta}))$ is non-linear and its approximate computation requires a m order polynomial expansion. The summary statistics are $\boldsymbol{t}(\boldsymbol{x}) = ([\boldsymbol{x}]^{\boldsymbol{k}})_{\boldsymbol{k}} \forall \boldsymbol{k} \in \mathbb{N}^d : \sum_j \boldsymbol{k}_j = m', m' \leq m$, where $[\boldsymbol{x}]^{\boldsymbol{k}} = \prod_{j=1}^d \boldsymbol{x}_j^{\boldsymbol{k}_j}$. An example of a second order (i.e. $m = 2$) approximate summary statistic $\boldsymbol{t}(\boldsymbol{x})$ for logistic regression when $d = 4$ is given below.

$$\boldsymbol{t}(\boldsymbol{x}) = \Big[1, x_1, x_2, x_3, x_4, x_1^2, x_2^2, x_3^2, x_4^2, x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4\Big] \tag{4}$$

The entries for $\boldsymbol{\mu}_{\boldsymbol{s}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{s}}$ for the second part of the summand is given by the following.

1. $\mathbb{E}[x_i^a x_j^b] = \left\{ \begin{array}{ll} \mathbb{E}_{\boldsymbol{x}}[x_i^a x_j^b] & \text{for even a+b's} \\ 0 & \text{for odd a+b's} \end{array} \right\}, a, b : a + b = m' \leq m$
2. $\text{Cov}[x_i^a x_j^b, x_k^c x_l^d] = \mathbb{E}_{\boldsymbol{x}}[x_i^a x_j^b x_k^c x_l^d] - \mathbb{E}[x_i^a x_j^b]\mathbb{E}[x_k^c x_l^d] a, b : a + b = m', c + d = m', m' \leq m$.

According to the Isserlis' theorem, only the even degree moments are non zero. Therefore we have the following cases.

$$\text{Cov}[x_i^a x_j^b, x_k^c x_l^d] = \left\{ \begin{array}{ll} \mathbb{E}_{\boldsymbol{x}}[x_i^a x_j^b x_k^c x_l^d] - \mathbb{E}_{\boldsymbol{x}}[x_i^a x_j^b]\mathbb{E}_{\boldsymbol{x}}[x_k^c x_l^d] & \text{a+b and c+d are even} \\ \mathbb{E}_{\boldsymbol{x}}[x_i^a x_j^b x_k^c x_l^d] & \text{a+b and c+d are odd} \\ 0 & \text{a+b+c+d is odd} \end{array} \right\}$$

2.2 $y \log(\log(1 + \exp(\boldsymbol{x}^T\boldsymbol{\theta})))$

The entries of $\boldsymbol{t}(\boldsymbol{x})$ for $\log(\log(1 + \exp(\boldsymbol{x}^T\boldsymbol{\theta})))$ are the same as those are for $\log(1 + \exp(\boldsymbol{x}^T\boldsymbol{\theta}))$. So let's compute the entries for $\boldsymbol{\mu_s}$ for $y\boldsymbol{t}(\boldsymbol{x})$.

$$\mathbb{E}_{\boldsymbol{x}}[\boldsymbol{t}(\boldsymbol{x})y] = \mathbb{E}_{\boldsymbol{x}}[x_i^a x_j^b \mathbb{E}_{y|\boldsymbol{x}}[y]] = \mathbb{E}_{\boldsymbol{x}}[x_i^a x_j^b \log(1 + \exp(\boldsymbol{x}^T\boldsymbol{\theta}))] \approx \mathbb{E}_{\boldsymbol{x}}\Big[x_i^a x_j^b \Big(\log(2) + \frac{\boldsymbol{x}^T\boldsymbol{\theta}}{2} + \frac{(\boldsymbol{x}^T\boldsymbol{\theta})^2}{8} - \frac{(\boldsymbol{x}^T\boldsymbol{\theta})^4}{192}\Big)\Big].$$

Here $\log(1+\exp(\boldsymbol{x}^T\boldsymbol{\theta}))$ is approximated using the first four terms of the its Taylor expansion. Each monomial in the expansion of $(\boldsymbol{x}^T\boldsymbol{\theta})^p, p \in \mathbb{N}^{\geq 0}$ has degree p and we expand these monomials using the multinomial theorem. Once again after applying the Isserlis' theorem, we have the following cases.

$$\mathbb{E}[x_i^a x_j^b \log(1 + \exp(\boldsymbol{x}^T\boldsymbol{\theta}))] = \left\{ \begin{array}{ll} \mathbb{E}_{\boldsymbol{x}}[x_i^a x_j^b \Big(\log(2) + \frac{(\boldsymbol{x}^T\boldsymbol{\theta})^2}{8} - \frac{(\boldsymbol{x}^T\boldsymbol{\theta})^4}{192}\Big)] & \text{for even a+b's} \\ \mathbb{E}_{\boldsymbol{x}}[x_i^a x_j^b \Big(\frac{\boldsymbol{x}^T\boldsymbol{\theta}}{2}\Big)] & \text{for odd a+b's} \end{array} \right\}, a, b : a + b = m' \leq m. \tag{5}$$

Similarly the terms of $\boldsymbol{\Sigma_s}$ for all $a, b : a + b = m' \leq m, c + d = m', m' \leq m$ are:

$$\begin{aligned} \text{Cov}[x_i^a x_j^b y, x_k^c x_l^d y] &= \mathbb{E}_{\boldsymbol{x}}[x_i^a x_j^b x_k^c x_l^d \mathbb{E}_{y|\boldsymbol{x}}[y^2]] - \mathbb{E}_{\boldsymbol{x}}[x_i^a x_j^b \mathbb{E}_{y|\boldsymbol{x}}[y]]\mathbb{E}_{\boldsymbol{x}}[x_k^c x_l^d \mathbb{E}_{y|\boldsymbol{x}}[y]] \\ &= \mathbb{E}_{\boldsymbol{x}}[x_i^a x_j^b x_k^c x_l^d \log^2(1 + \exp(\boldsymbol{x}^T\boldsymbol{\theta}))] - \mathbb{E}_{\boldsymbol{x}}[x_i^a x_j^b \log(1 + \exp(\boldsymbol{x}^T\boldsymbol{\theta}))]\mathbb{E}_{\boldsymbol{x}}[x_k^c x_l^d \log(1 + \exp(\boldsymbol{x}^T\boldsymbol{\theta}))]. \end{aligned}$$

The second part of this subtraction can be evaluated using Equation 5. Next, we evaluate $\mathbb{E}_{\boldsymbol{x}}[x_i^a x_j^b x_k^c x_l^d \log^2(1 + \exp(\boldsymbol{x}^T\boldsymbol{\theta}))]$. Using the Taylor series expansion,

$$\log^2(1 + \exp(\boldsymbol{x}^T\boldsymbol{\theta})) \approx \log^2(2) + (\boldsymbol{x}^T\boldsymbol{\theta})\log(2) + \frac{(\boldsymbol{x}^T\boldsymbol{\theta})^2(1 + \log(2))}{4} + \frac{(\boldsymbol{x}^T\boldsymbol{\theta})^3}{8}.$$

The surviving even degree moments that we evaluate are.

$$\mathbb{E}[x_i^a x_j^b x_k^c x_l^d \log^2(1 + \exp(\boldsymbol{x}^T\boldsymbol{\theta}))] = \left\{ \begin{array}{ll} \mathbb{E}_{\boldsymbol{x}}\Big[x_i^a x_j^b x_k^c x_l^d \Big(\log^2(2) + \frac{(\boldsymbol{x}^T\boldsymbol{\theta})^2(1+\log(2))}{4}\Big)\Big] & \text{for even a+b+c+d's} \\ \mathbb{E}_{\boldsymbol{x}}\Big[x_i^a x_j^b x_k^c x_l^d \Big((\boldsymbol{x}^T\boldsymbol{\theta})\log(2) + \frac{(\boldsymbol{x}^T\boldsymbol{\theta})^3}{8}\Big)\Big] & \text{for odd a+b+c+d's} \end{array} \right\}.$$

These expressions are further simplified using the multinomial theorem once again.

3 Sensitivity results

3.1 Individual sensitivities

Lemma 3.1.1. *Consider two vectors $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^d$ such that $||\boldsymbol{x}||_2 \leq R$ and $||\boldsymbol{x}'||_2 \leq R$. Then, an elementary analysis shows that*

$$\begin{aligned} ||t_1(\boldsymbol{x}) - t_1(\boldsymbol{x}')||_2 &\leq 2R, \\ ||t_2(\boldsymbol{x}) - t_2(\boldsymbol{x}')||_2 &\leq \sqrt{2}R^2. \end{aligned}$$

When considering the Gaussian mechanism for releasing $t_1(\boldsymbol{x})$ or $t_2(\boldsymbol{x})$ such that element-wise Gaussian noises of variances $4R^2\sigma^2$ and $2R^4\sigma^2$ are added to $t_1(\boldsymbol{x})$ and $t_2(\boldsymbol{x})$, respectively, their (ε, δ)-analyses are equivalent to the analysis of Gaussian mechanism with sensitivity 1 and variance σ^2. However, when releasing the linear and quadratic terms simultaneously, a better utility can be obtained. To this end, we need the following relations.

3.2 Second order terms

In the case of $m = 2$, we release the terms $(x_i^2, x_i x_j)$ with multipliers $(1, \sqrt{2})$. By rearranging, we have

$$\begin{aligned}
\|t_2(x) - t_2(x')\|_2^2 &= \sum_{i=1}^{d}(x_i^2 - x_i'^2)^2 + 2\sum_{i>j}(x_i x_j - x_i' x_j')^2 \\
&= \sum_i x_i^4 + \sum_i x_i'^4 - 2\sum_i x_i^2 x_i'^2 + 2\sum_{i>j} x_i^2 x_j^2 + 2\sum_{i>j} x_i'^2 x_j'^2 - 4\sum_{i>j} x_i x_j x_i' x_j' \\
&= \left(\sum_i x_i^4 + 2\sum_{i>j} x_i^2 x_j^2\right) + \left(\sum_i x_i'^4 + 2\sum_{i>j} x_i'^2 x_j'^2\right) - 2\sum_i x_i^2 x_i'^2 - 4\sum_{i>j} x_i x_j x_i' x_j' \\
&= \left(\sum_i x_i^4 + \sum_{i\neq j} x_i^2 x_j^2\right) + \left(\sum_i x_i'^4 + \sum_{i\neq j} x_i'^2 x_j'^2\right) - 2\sum_i x_i^2 x_i'^2 - 2\sum_{i\neq j} x_i x_j x_i' x_j' \\
&= \|x\|_2^4 + \|x'\|_2^4 - 2\langle x, x'\rangle^2.
\end{aligned}$$

3.3 Third and fourth order terms

We first illustrate the general case with the cases $m = 3$ and $m = 4$. In the next section we describe the release mechanism and give its tight sensitivity for general m.

In the case of $m = 3$, we release all the distinct terms of the form

$$(x_i^3, x_i^2 x_j, x_i x_j x_k)$$

with the corresponding multipliers $(1, \sqrt{3}, \sqrt{3})$. By rearranging, we have

$$\begin{aligned}
\|t_3(x) - t_3(x')\|_2^2 &= \sum_{i=1}^{d}(x_i^3 - x_i'^3)^2 + 3\sum_{i\neq j}(x_i^2 x_j - x_i'^2 x_j')^2 + 3\sum_{i>j>k}(x_i x_j x_k - x_i' x_j' x_k')^2 \\
&= \sum_i x_i^6 + \sum_i x_i'^6 - 2\sum_i x_i^3 x_i'^3 + 3\sum_{i\neq j} x_i^4 x_j^2 + 3\sum_{i\neq j} x_i'^4 x_j'^2 \\
&\quad - 6\sum_{i\neq j} x_i^2 x_j x_i'^2 x_j' + 3\sum_{i>j>k} x_i^2 x_j^2 x_k^2 + 3\sum_{i>j>k} x_i'^2 x_j'^2 x_k'^2 - 6\sum_{i>j>k} x_i x_j x_k x_i' x_j' x_k' \\
&= \left(\sum_i x_i^6 + 3\sum_{i\neq j} x_i^4 x_j^2 + 3\sum_{i>j>k} x_i^2 x_j^2 x_k^2\right) + \left(\sum_i x_i'^6 + 3\sum_{i\neq j} x_i'^4 x_j'^2 + 3\sum_{i>j>k} x_i'^2 x_j'^2 x_k'^2\right) \\
&\quad - 2\sum_i x_i^3 x_i'^3 - 6\sum_{i\neq j} x_i^2 x_j x_i'^2 x_j' - 6\sum_{i>j>k} x_i x_j x_k x_i' x_j' x_k' \\
&= \|x\|_2^6 + \|x'\|_2^6 - 2\langle x, x'\rangle^3.
\end{aligned}$$

In the case of $m = 4$, we release all the distinct the terms of the forms

$$(x_i^4, x_i^3 x_j, x_i^2 x_j^2, x_i^2 x_j x_k, x_i x_j x_k x_\ell)$$

with the corresponding multipliers $(\sqrt{4}, \sqrt{6}, \sqrt{6}, \sqrt{4})$. By rearranging, we have

$$
\begin{aligned}
\|t_4(x)-t_4(x')\|_2^2 &= \sum_{i=1}^{d}(x_i^4-x_i'^4)^2+4\sum_{i\neq j}(x_i^3x_j-x_i'^3x_j')^2+6\sum_{i>j}(x_i^2x_j^2-x_i'^2x_j'^2)^2\\
&+6\sum_{i\neq j,i\neq k,j\neq k}(x_i^2x_jx_k-x_i'^2x_j'x_k')^2+4\sum_{i>j>k>\ell}(x_ix_jx_kx_\ell-x_i'x_j'x_k'x_\ell')^2\\
&=\Bigg(\sum_i x_i^8+\sum_i x_i'^8-2\sum_i x_i^4x_i'^4\Bigg)+\Bigg(4\sum_{i\neq j}x_i^6x_j^2+4\sum_{i\neq j}x_i'^6x_j'^2\\
&-8\sum_{i\neq j}x_i^3x_jx_i'^3x_j'\Bigg)+\Bigg(6\sum_{i>j}x_i^4x_j^4+6\sum_{i>j}x_i'^4x_j'^4-12\sum_{i>j}x_i^2x_j^2x_i'^2x_j'^2\Bigg)\\
&+\Bigg(6\sum_{i\neq j,i\neq k,j\neq k}x_i^4x_j^2x_k^2+6\sum_{i\neq j,i\neq k,j\neq k}x_i'^4x_j'^2x_k'^2-12\sum_{i\neq j,i\neq k,j\neq k}x_i^2x_jx_kx_i'^2x_j'x_k'\Bigg)\\
&+\Bigg(4\sum_{i>j>k>\ell}x_i^2x_j^2x_k^2x_\ell^2+4\sum_{i>j>k>\ell}x_i'^2x_j'^2x_k'^2x_\ell'^2-8\sum_{i>j>k>\ell}x_ix_jx_kx_\ell x_i'x_j'x_k'x_\ell'\Bigg)\\
&=\Bigg(\sum_i x_i^8+4\sum_{i\neq j}x_i^6x_j^2+6\sum_{i>j}x_i^4x_j^4+4\sum_{i>j>k>\ell}x_i^2x_j^2x_k^2x_\ell^2\Bigg)+\\
&+\Bigg(\sum_i x_i'^8+4\sum_{i\neq j}x_i'^6x_j'^2+6\sum_{i>j}x_i'^4x_j'^4+4\sum_{i>j>k>\ell}x_i'^2x_j'^2x_k'^2x_\ell'^2\Bigg)\\
&-2\Bigg(\sum_i x_i^4x_i'^4+4\sum_{i\neq j}x_i^3x_jx_i'^3x_j'+6\sum_{i\neq j,i\neq k,j\neq k}x_i^2x_jx_kx_i'^2x_j'x_k'\\
&+6\sum_{i>j}x_i^2x_j^2x_i'^2x_j'^2+4\sum_{i>j>k>\ell}x_ix_jx_kx_\ell x_i'x_j'x_k'x_\ell'\Bigg)\\
&=\|x\|_2^8+\|x'\|_2^8-2\langle x,x'\rangle^4.
\end{aligned}
$$

3.4 General case

For a general m, if we release each distinct mth order termof the form

$$x_{i_1}^{k_1}\cdots x_{i_{m'}}^{k_{m'}},$$

where $\sum_i k_i = m$, $1\le m'\le m$, multiplied with the multinomial factor $\sqrt{\binom{m}{k_1,\ldots,k_{m'}}}$, then the function $t_m(x)$ has the sensitivity

$$\|t_m(x)-t_m(x')\|_2^2=\|x\|_2^{2m}+\|x'\|_2^{2m}-2\langle x,x'\rangle^m.$$

This is shown similarly as above for $t_2(x), t_3(x)$ and $t_4(x)$. Namely, we have

$$
\begin{aligned}
\|t_m(x)-t_m(x')\|_2^2 &= \sum_{k_1+\ldots k_{m'}=m,\,1\le m'\le m}\ \sum_{i_1,\ldots i_{m'}}\binom{m}{k_1,\ldots,k_{m'}}(x_{i_1}^{k_1}\cdots x_{i_{m'}}^{k_{m'}}-x_{i_1}'^{k_1}\cdots x_{i_{m'}}'^{k_{m'}})^2\\
&=\Bigg(\sum_{k_1+\ldots k_{m'}=m,\,1\le m'\le m}\ \sum_{i_1,\ldots i_{m'}}\binom{m}{k_1,\ldots,k_{m'}}x_{i_1}^{2k_1}\cdots x_{i_{m'}}^{2k_{m'}}\Bigg)\\
&+\Bigg(\sum_{k_1+\ldots k_{m'}=m,\,1\le m'\le m}\ \sum_{i_1,\ldots i_{m'}}\binom{m}{k_1,\ldots,k_{m'}}x_{i_1}'^{2k_1}\cdots x_{i_{m'}}'^{2k_{m'}}\Bigg)\\
&-2\cdot\Bigg(\sum_{k_1+\ldots k_{m'}=m,\,1\le m'\le m}\ \sum_{i_1,\ldots i_{m'}}\binom{m}{k_1,\ldots,k_{m'}}x_{i_1}^{k_1}x_{i_1}'^{k_1}\cdots x_{i_{m'}}^{k_{m'}}x_{i_{m'}}'^{k_{m'}}\Bigg)\\
&=\|x\|_2^{2m}+\|x'\|_2^{2m}-2\langle x,x'\rangle^m,
\end{aligned}
$$

where $\sum_{k_1+\ldots k_{m'}=m,\, 1\leq m'\leq m}$ denotes a sum over all *combinations* of positive integers $(k_1, \ldots, k_{m'})$ such that $k_1 + \ldots k_{m'} = m$ and

$$\sum_{i_1,\ldots i_{m'}} x_{i_1}^{k_1} \cdots x_{i_{m'}}^{k_{m'}}$$

denotes a sum over all different monomials with $(k_1, \ldots, k_{m'})$ as exponents.

4 Preliminary experiments on Poisson regression

Implementation. We used Metropolis-Hastings algorithm to infer the model parameter posteriors for Poisson regression. We gave the regression coefficients $\boldsymbol{\theta_s}$ a standard normal prior and the data covariance $\boldsymbol{\Sigma_s}$ an Inverse Wishart prior. As mentioned in the main draft, to the best of our knowledge, this is the first work that analyzes Poisson regression under DP constraints. Instead of employing the Isserlis theorem, we approximated the normal approximation parameters using MC integration. Specifying a more accurate, efficient, and scalable model for Poisson regression in sophisticated probabilistic programmings frameworks such as Stan is marked as a future exercise. We use synthetic data of 500 samples generated with $\boldsymbol{\theta} = [0.3, -0.6, 0.8]$ and a valid non-identity co-variance matrix. We filter out $||\boldsymbol{x}||_2 > R_x = 1$. The proposal standard deviation for the MH sampler was set to 0.01. Our sampler runs for 50,000 iterations, out of which we discard the first 25,000 burn-in samples. We repeat each inference for 5 times.

Results. Figure 1 compares private and non-private empirical CDFs for $\boldsymbol{\theta}'$s for various ϵ values within range $[0.1, 1.1]$. The last plot shows the Kolmogorov-Smirnov scores between these CDFs for a few ϵ values in the same range. We note that the private CDFs tend to (partially) overlap on their non-private variants as ϵ increases. We suspect that the overlap is not as strong as it is in logistic regression due to a) more noise, which is a consequence of significantly more number of approximate sufficient statistics and larger range of y (already explained by Lemma 3.5 in the main draft), b) smaller sample size of 500 and smaller number of inference repeats causing more uncertainty.

We believe that it may be possible to improve these results with faster converging sampling algorithms and by designing better prior distributions. However, these preliminary explorations do demonstrate the merit of our model and tight sensitivity resultst.

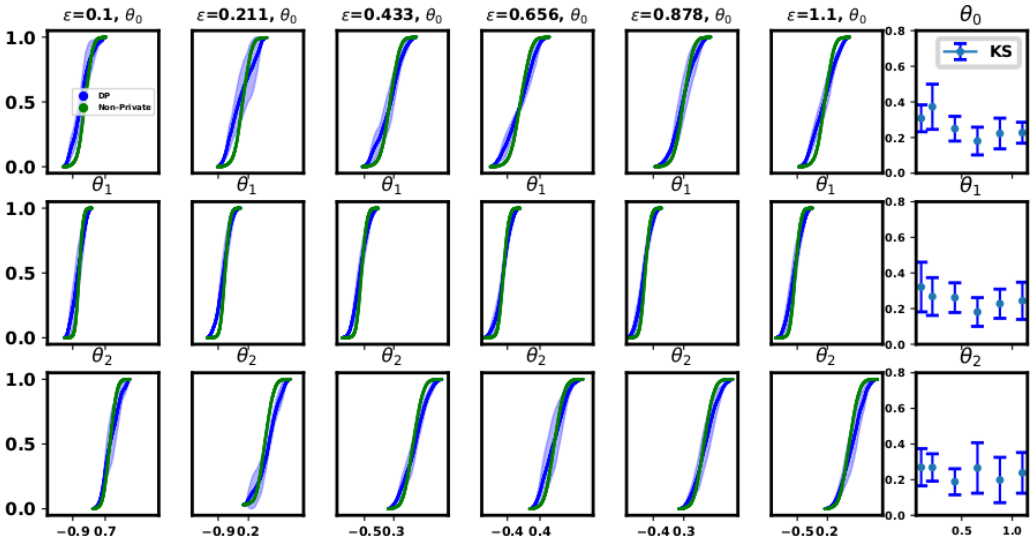

Figure 1: Comparison of differentially private and non-private empirical CDFs for $\boldsymbol{\theta}$'s posteriors for Poisson regression for various ϵ values. We use a synthetic dataset of $N = 500$ samples and $R_x = 1, R_y = 5, \delta = 10^{-5}$. The right-most column shows the Kolmogorov-Smirnov scores between non-private and private empirical CDFs for the same set of ϵ values.

References

[1] G. C. Wick. The evaluation of the collision matrix. *Phys. Rev.*, 80:268–272, Oct 1950.