APPENDIX TO: DIFFERENTIALLY PRIVATE CONFIDENCE INTERVALS FOR EMPIRICAL RISK MINIMIZATION

YUE WANG, DANIEL KIFER, AND JAEWOO LEE

The main manuscript is available at https://doi.org/10.29012/jpc.660.

APPENDIX B. COMPLETE EXPERIMENTAL RESULTS

B.1. **Allocation for the Privacy Budget.** See Figures 1 to 12.

B.2. **Empirical Sample Complexity of Private Confidence Intervals.** See Figures 13 to 20.

*Key words and phrases:* Differential Privacy, Objective Perturbation, Output Perturbation, Confidence Intervals.
Figure 1: [\(\epsilon\text{-DP, objective perturbation, logistic regression}\)] Coverage percentage and average length of confidence intervals vs. \(\epsilon_1/\epsilon\) for objective perturbation based \(\epsilon\text{-DP confidence intervals}\) for linear regression with a total privacy budget of \(\epsilon = 1.0\). \(\epsilon_2 = \epsilon_3 = (\epsilon - \epsilon_1)/2\), \(c = 0.001\).
Figure 2: \([\epsilon\text{-DP, objective perturbation, SVM}]\) Coverage percentage and average length of confidence intervals vs. \(\epsilon_1/\epsilon\) for objective perturbation based \(\epsilon\text{-DP} \) confidence intervals for SVM with a total privacy budget of \(\epsilon = 1.0\). \(\epsilon_2 = \epsilon_3 = (\epsilon - \epsilon_1)/2, c = 0.001, h = 1.0.\)
Figure 3: [zCDP, objective perturbation, logistic regression] Coverage percentage and average length of confidence intervals vs. $\rho_1/\rho$ for objective perturbation based zCDP confidence intervals for linear regression with a total privacy budget of $\rho = 0.5$. $\rho_2 = \rho_3 = (\rho - \rho_1)/2$, $c = 0.001$. 
<table>
<thead>
<tr>
<th>Coverage Percentage</th>
<th>CI Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.90</td>
</tr>
<tr>
<td>0.50</td>
<td>0.92</td>
</tr>
<tr>
<td>0.75</td>
<td>0.94</td>
</tr>
<tr>
<td>1.00</td>
<td>0.96</td>
</tr>
</tbody>
</table>

- (a) Adult, n = 5,000, d = 1
- (b) Adult, n = 5,000, d = 5
- (c) Adult, n = 10,000, d = 10
- (d) US, n = 5,000, d = 1
- (e) US, n = 5,000, d = 5
- (f) US, n = 10,000, d = 10

Figure 4: [zCDP, objective perturbation, SVM] Coverage percentage and average length of confidence intervals vs. $\rho_1/\rho$ for objective perturbation based zCDP confidence intervals for SVM with a total privacy budget of $\rho = 0.5$. $\rho_2 = \rho_3 = (\rho - \rho_1)/2$, $c = 0.001$, $h = 1.0$. 
Figure 5: [\(\epsilon\)-DP, output perturbation, logistic regression] Coverage percentage and average length of confidence intervals vs. \(\epsilon_1/\epsilon\) for output perturbation based \(\epsilon\)-DP confidence intervals for linear regression with a total privacy budget of \(\epsilon = 1.0\). \(\epsilon_2 = \epsilon_3 = (\epsilon - \epsilon_1)/2\), \(c = 0.001\).
Figure 6: [\(\epsilon\text{-DP, output perturbation, SVM}\)] Coverage percentage and average length of confidence intervals vs. \(\epsilon_1/\epsilon\) for output perturbation based \(\epsilon\)-DP confidence intervals for SVM with a total privacy budget of \(\epsilon = 1.0\). \(\epsilon_2 = \epsilon_3 = (\epsilon - \epsilon_1)/2\), \(c = 0.001\), \(h = 1.0\).
Figure 7: [zCDP, output perturbation, logistic regression] Coverage percentage and average length of confidence intervals vs. $\rho_1/\rho$ for output perturbation based zCDP confidence intervals for linear regression with a total privacy budget of $\rho = 0.5$. $\rho_2 = \rho_3 = (\rho - \rho_1)/2$, $c = 0.001$. 

(A) KDDCUP99, $n = 5,000$, $d = 1$

(B) KDDCUP99, $n = 5,000$, $d = 5$

(C) KDDCUP99, $n = 10,000$, $d = 10$

(D) Banking, $n = 5,000$, $d = 1$

(E) Banking, $n = 5,000$, $d = 5$

(F) Banking, $n = 10,000$, $d = 10$
Figure 8: [zCDP, output perturbation, SVM] Coverage percentage and average length of confidence intervals vs. $\rho_1/\rho$ for output perturbation based zCDP confidence intervals for SVM with a total privacy budget of $\rho = 0.5$. $\rho_2 = \rho_3 = (\rho - \rho_1)/2$, $c = 0.001$, $h = 1.0$. 

(a) Adult, $n = 5,000$, $d = 1$

(b) Adult, $n = 5,000$, $d = 5$

(c) Adult, $n = 10,000$, $d = 10$

(d) KDDCUP99, $n = 5,000$, $d = 1$

(e) KDDCUP99, $n = 5,000$, $d = 5$

(f) KDDCUP99, $n = 10,000$, $d = 10$
Figure 9: [ε-DP, objective perturbation] Coverage percentage and average length of confidence intervals vs. $\varepsilon_2/(\varepsilon - \varepsilon_1)$ for objective perturbation based ε-DP confidence intervals with a total privacy budget of $\varepsilon = 1.0$. $\varepsilon_1 = 0.65$, $c = 0.001$, $h = 1.0$. 

(A) Adult, $n = 5,000$, $d = 1$, logistic regression

(B) Adult, $n = 5,000$, $d = 5$, logistic regression

(C) Adult, $n = 10,000$, $d = 10$, logistic regression

(D) Banking, $n = 5,000$, $d = 1$, SVM

(E) Banking, $n = 5,000$, $d = 5$, SVM

(F) Banking, $n = 10,000$, $d = 10$, SVM
Figure 10: [zCDP, objective perturbation] Coverage percentage and average length of confidence intervals vs. $\rho_2/(\rho - \rho_1)$ for objective perturbation based zCDP confidence intervals with a total privacy budget of $\rho = 0.5$. $\rho_1 = 0.45$, $c = 0.001$, $h = 1.0$. 

(A) Banking, $n = 5,000$, $d = 1$, logistic regression

(B) Banking, $n = 5,000$, $d = 5$, logistic regression

(C) Banking, $n = 10,000$, $d = 10$, logistic regression

(D) Adult, $n = 5,000$, $d = 1$, SVM

(E) Adult, $n = 5,000$, $d = 5$, SVM

(F) Adult, $n = 10,000$, $d = 10$, SVM
Figure 11: [\epsilon\text{-DP, output perturbation}] Coverage percentage and average length of confidence intervals vs. $\epsilon_2/(\epsilon - \epsilon_1)$ for output perturbation based $\epsilon$-DP confidence intervals with a total privacy budget of $\epsilon = 1.0$. $\epsilon_1 = 0.8$, $c = 0.001$, $h = 1.0$. 

(a) BR, $n = 5,000$, $d = 1$, logistic regression
(b) BR, $n = 5,000$, $d = 5$, logistic regression
(c) BR, $n = 10,000$, $d = 10$, logistic regression
(d) US, $n = 5,000$, $d = 1$, SVM
(e) US, $n = 5,000$, $d = 5$, SVM
(f) US, $n = 10,000$, $d = 10$, SVM
Figure 12: \([\text{zCDP, output perturbation}]\) Coverage percentage and average length of confidence intervals vs. \(\rho_2/(\rho - \rho_1)\) for output perturbation based zCDP confidence intervals with a total privacy budget of \(\rho = 0.5\). \(\rho_1 = 0.45\), \(c = 0.001\), \(h = 1.0\).
Figure 13: Coverage percentage from 1000 confidence intervals as a function of the sample size \( n \) and the dimensionality \( d \) on Adult dataset for logistic regression. \( \epsilon = 1.0, \rho = \epsilon^2/2 = 0.5, c = 0.001 \).

Figure 14: Average length from 1000 confidence intervals as a function of the sample size \( n \) and the dimensionality \( d \) on Adult dataset for logistic regression. \( \epsilon = 1.0, \rho = \epsilon^2/2 = 0.5, c = 0.001 \).
Figure 15: Coverage percentage from 1000 confidence intervals as a function of the sample size \( n \) and dimensionality \( d \) on Banking dataset for SVM. \( \epsilon = 1.0, \rho = \epsilon^2/2 = 0.5, c = 0.001, h = 1.0 \).

Figure 16: Average length from 1000 confidence intervals as a function of the sample size \( n \) and dimensionality \( d \) on Banking dataset for SVM. \( \epsilon = 1.0, \rho = \epsilon^2/2 = 0.5, c = 0.001, h = 1.0 \).
Figure 17: Coverage percentage from 1000 confidence intervals as a function of the sample size $n$ and the total privacy budget $\epsilon$ (or $\rho$ where $\rho = \epsilon^2/2$) on KDDCUP99 dataset for logistic regression. $d = 10, c = 0.001$.

Figure 18: Average length from 1000 confidence intervals as a function of the sample size $n$ and the total privacy budget $\epsilon$ (or $\rho$ where $\rho = \epsilon^2/2$) on KDDCUP99 dataset for logistic regression. $d = 10, c = 0.001$. 
Figure 19: Coverage percentage from 1000 confidence intervals as a function of the sample size $n$ and the total privacy budget $\epsilon$ (or $\rho$ where $\rho = \epsilon^2/2$) on US dataset for SVM. $d = 10$, $c = 0.001$, $h = 1.0$.

**B.3. The Overhead on Sample Complexity for Differential Privacy.** See Figures 21 and 22.

**B.4. Comparison among the Private Confidence Intervals and the Variability Intervals.** See Figures 23 to 28.

**B.5. Modeling the Relationship between Length of the Intervals and Other Parameters.** See Figures 29 to 34.
Figure 20: Average length from 1000 confidence intervals as a function of the sample size $n$ and the total privacy budget $\epsilon$ (or $\rho$ where $\rho = \epsilon^2/2$) on US dataset for SVM. $d = 10$, $c = 0.001$, $h = 1.0$.

Figure 21: The mapping between sample complexities such that the average length of the non-private confidence intervals is equivalent to that of the private confidence intervals for logistic regression. $\epsilon = 1.0$ (corresponds to $\rho = 0.5$), $d = 10$, $c = 0.001$. 
APPENDIX TO: DIFFERENTIALLY PRIVATE CONFIDENCE INTERVALS

Figure 22: The mapping between sample complexities such that the average length of the non-private confidence intervals is equivalent to that of the private confidence intervals for SVM. \( \epsilon = 1.0 \) (corresponds to \( \rho = 0.5 \)), \( d = 10 \), \( c = 0.001 \), \( h = 1.0 \).

Figure 23: Comparison among length of intervals with varying \( n \) for logistic regression on Adult dataset. \( d = 10 \), \( \epsilon = 1.0 \) (corresponds to \( \rho = 0.5 \)), \( c = 0.001 \).

Figure 24: Comparison among length of intervals with varying \( n \) for SVM on Banking dataset. \( d = 10 \), \( \epsilon = 1.0 \) (corresponds to \( \rho = 0.5 \)), \( c = 0.001 \), \( h = 1.0 \).
Figure 25: Comparison among length of intervals with varying $d$ for logistic regression on Adult dataset. $n = 15,000$, $\epsilon = 1.0$ (corresponds to $\rho = 0.5$), $c = 0.001$.

Figure 26: Comparison among length of intervals with varying $d$ for SVM on Banking dataset. $n = 15,000$, $\epsilon = 1.0$ (corresponds to $\rho = 0.5$), $c = 0.001$, $h = 1.0$.

Figure 27: Comparison among length of intervals with varying $\epsilon$ (or $\rho$ with $\rho = \epsilon^2/2$) for logistic regression on KDDCUP99 dataset. $n = 15,000$, $d = 10$, $c = 0.001$. 
Figure 28: Comparison among length of intervals with varying \( \epsilon \) (or \( \rho \) with \( \rho = \epsilon^2/2 \)) for SVM on US dataset. \( n = 15,000, d = 10, c = 0.001, h = 1.0 \).

Figure 29: Relationship between average length of the confidence intervals and the sample size \( n \) for logistic regression. \( d = 5, c = 0.001, \epsilon = 1.0 \) (corresponds to \( \rho = 0.5 \)). The fitted curve is \( \frac{c_0}{n} \frac{c_1}{\sqrt{n}} \).
Figure 30: Relationship between average length of the confidence intervals and the sample size \( n \) for logistic regression. \( d = 5, c = 0.001, \epsilon = 1.0 \) (corresponds to \( \rho = 0.5 \)). The fitted curve is \( \frac{c}{\sqrt{n}} \), that shows the length for the privacy-preserving confidence intervals is not proportional to \( \frac{1}{\sqrt{n}} \) as in the non-private case.
Figure 31: Relationship between average length of the confidence intervals and the sample size $n$ for SVM. $d = 5$, $c = 0.001$, $\epsilon = 1.0$ (corresponds to $\rho = 0.5$), $h = 1.0$. The fitted curve is $\frac{c_0}{n} + \frac{c_1}{\sqrt{n}}$. 
Figure 32: Relationship between average length of the confidence intervals and the sample size $n$ for SVM. $d = 5$, $c = 0.001$, $\epsilon = 1.0$ (corresponds to $\rho = 0.5$), $h = 1.0$. The fitted curve is $\frac{c}{\sqrt{n}}$, that shows the length for the privacy-preserving confidence intervals is not proportional to $\frac{1}{\sqrt{n}}$ as in the non-private case.
Figure 33: Relationship between average length of the confidence intervals and the total privacy budget $\epsilon$ (or $\rho = \epsilon^2/2$) for logistic regression. $n = 10000$, $d = 5$, $c = 0.001$. The fitted curve is $\sqrt{\frac{2c}{\epsilon}} + c_1$ for $\epsilon$-DP, $\sqrt{\frac{2\rho}{\rho}} + c_1$ for zCDP.
Figure 34: Relationship between average length of the confidence intervals and the total privacy budget $\epsilon$ (or $\rho = \epsilon^2/2$) for SVM. $n = 10000$, $d = 5$, $c = 0.001$, $h = 1.0$. The fitted curve is $\sqrt{\frac{\epsilon^2}{c^2}} + c_1$ for $\epsilon$-DP, $\sqrt{\frac{\epsilon^2}{\rho}} + c_1$ for zCDP.