Differential Privacy via DRO: GitHub Appendices

Aras Selvi*[1], Huikang Liu[2] and Wolfram Wiesemann[1]

[1]Imperial College Business School, Imperial College London, United Kingdom
[2]Research Institute for Interdisciplinary Sciences, School of Information Management and Engineering, Shanghai University of Finance and Economics, China

April 27, 2023

Suboptimalities of Several Mainstream Assumptions

In the main paper we claimed that several mainstream assumptions taken in the literature, which we do not take, are **not** without loss of generality. Indeed, we next show some counterexamples of each popular assumption on optimal noise distributions: *(i)* they are monotone around the origin; *(ii)* they are symmetric around the origin; *(iii)* they come from a certain family of distributions.

Monotone Distributions

To prove that monotone distributions are not always optimal, it is sufficient to give one counter example. To this end, we take $\varepsilon = 3.0$ and $\delta = 0.3$ and optimize the ℓ_1-loss over a sufficiently large support. While the optimal distribution achieves a loss of 0.1705, the monotonicity-constrained optimization problem gives a lower bound of 0.1830 for a discretization granularity of $\beta = 0.02$. In other words, the optimal monotone distribution cannot achieve a loss better than 0.1830 for any granularity and support, whereas a non-monotone distribution achieves a better loss already for $\beta = 0.02$.

Symmetric Distributions

We optimize symmetric ($c(x) = |x|$) and asymmetric ($c(x) = |x| + \mathbb{1}[x > 0] \cdot |x|$) loss functions and share the result in Figure 1. One can observe that while the distribution minimizing the former loss function is symmetric, the distribution minimizing the latter loss function is not symmetric around origin or any pother point. We note that increasing the asymmetry of loss functions further (*e.g.*, increasing the slope of $x > 0$), or increasing the privacy regime (that makes the optimal distributions use larger supports, hence incurring larger losses when $x > 0$) makes the asymmetry of the optimal distributions more severe.

*Corresponding author: a.selvi19@imperial.ac.uk

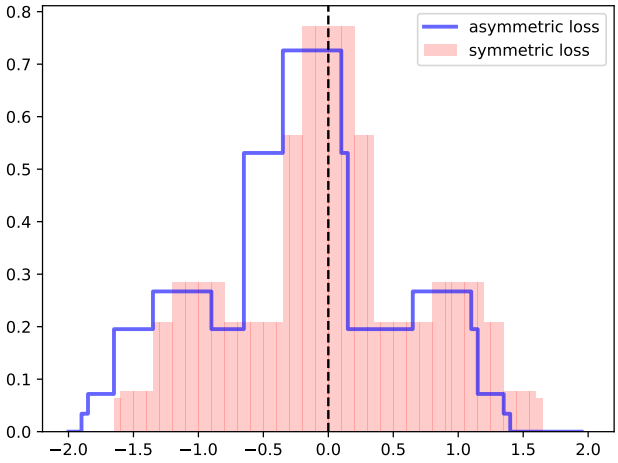

Figure 1: *Optimization-based noise distributions for synthetic data independent instances with* $\Delta f = 1$, $\beta = 0.05$, $\varepsilon = 1$, $\delta = 0.2$, *and two loss functions. The distribution minimizing the symmetric loss* ($c(x) = |x|$) *is shown in red shading, whereas the distribution minimizing the asymmetric loss* ($c(x) = |x| + \mathbb{1}[x > 0] \cdot |x|$) *is shown as blue lines.*

Restriction to a Family of Distributions

Finally, we give counterexamples on restricting the feasible noise distributions to a specific family. Although the previous counterexample on asymmetric loss functions would be sufficient for this purpose, we show a stronger result: even two different symmetric loss functions would yield the optimal solutions looks significantly different. In Figure 2, we observe that optimizing ℓ_1- and ℓ_2-losses result in distributions that could not belong to the same family of distributions, that is, there is no trivial density function that would generalize these distributions.

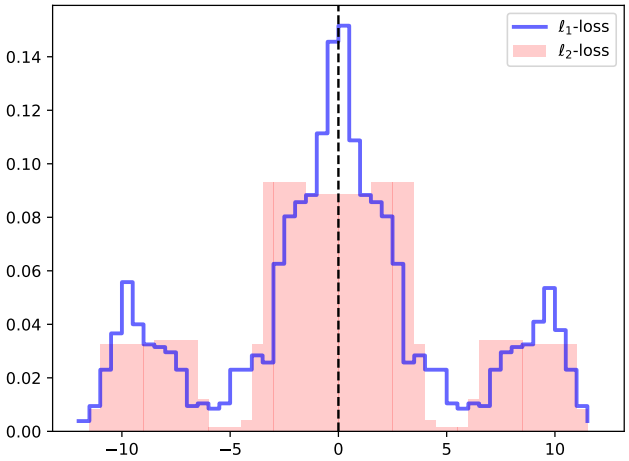

Figure 2: *Optimization-based noise distributions for synthetic data independent instances with* $\Delta f = 10$, $\beta = 0.5$, $\varepsilon = 1$, $\delta = 0.4$, *and two loss functions. The distribution minimizing the* ℓ_1*-loss* ($c(x) = |x|$) *is shown in red shading, whereas the distribution minimizing the* ℓ_2*-loss* ($c(x) = x^2$) *is shown as blue lines.*

Sampling Noise from Various Distributions

For the experiments on differentially private Naïve Bayes and proximal coordinate descent, we sample noise from various probability distributions to ensure privacy. Some distributions are easy to sample from by using the *Distributions* package of Julia, including the Laplace and Gaussian distributions. Here we give details on how to sample noise from the optimized distribution as well as the truncated Laplace distribution.

Optimized distribution.
Recall that, for data independent noise optimization, we solve an upper bound problem to obtain the mixture weights $\{p(j)\}_{j=1}^N$ of a mixture of uniform distributions. Recall that the probability of the noise being sampled from the j-th interval $\Pi_j(\beta)$ is $p(j)$. Hence, we first sample the interval from a discrete distribution with probabilities $\{p(j)\}_{j=1}^N$. Then, we sample the noise from a uniform distribution supported on $\Pi_j(\beta)$. Extension of this method to the data dependent setting is straightforward because although we have multiple optimized distributions, we sample noise from only the distribution corresponding to the true query value.

Truncated Laplace distribution.
For given $\delta \in (0, 0.5)$, $\varepsilon > 0$, $\Delta f > 0$, the Truncated Laplace distribution is defined by the probability density function:

$$f_{\text{TLap}}(x) = \begin{cases} B \cdot e^{\frac{-|x|}{\lambda}} & \text{for } x \in [-A, A] \\ 0, & \text{otherwise} \end{cases}$$

where $\lambda := \dfrac{\Delta f}{\varepsilon}$, $A := \dfrac{\Delta f}{\varepsilon} \cdot \log\left(1 + \dfrac{e^\varepsilon - 1}{2 \cdot \delta}\right)$, $B := \dfrac{1}{2 \cdot \lambda \cdot (1 - e^{-\frac{A}{\lambda}})}$. To sample noise from this distribution, we derive the inverse cumulative distribution function. To this end, we first derive the cumulative distribution function of f_{TLap}, which, after using some algebraic manipulations, can be expressed as:

$$F_{\text{TLap}}(x) = \begin{cases} 0 & \text{for } x \leq -A \\ 1 & \text{for } x \geq A \\ \dfrac{1}{2} - \text{sign}(x) \cdot \left[-\dfrac{1}{2} + B \cdot \lambda \cdot \exp(-|x|/\lambda) - B \cdot \lambda \cdot \exp(-A/\lambda)\right] & \text{for } x \in [-A, A]. \end{cases}$$

The inverse of this function can be obtained from the equation $F_{\text{TLap}}(F^{-1}_{\text{TLap}}(u)) = u$:

$$\begin{aligned} & \frac{1}{2} - \underbrace{\text{sign}(F^{-1}_{\text{TLap}}(u))}_{=\text{sign}(u-1/2)} \left[-\frac{1}{2} + B \cdot \lambda \cdot \exp\left(\frac{-|F^{-1}_{\text{TLap}}(u)|}{\lambda}\right) - B \cdot \lambda \cdot \exp\left(\frac{-A}{\lambda}\right)\right] = u \\ \iff & -\frac{1}{2} + B \cdot \lambda \cdot \exp\left(\frac{-|F^{-1}_{\text{TLap}}(u)|}{\lambda}\right) - B \cdot \lambda \cdot \exp\left(\frac{-A}{\lambda}\right) = \underbrace{\frac{u - 1/2}{-\text{sign}(u - 1/2)}}_{=-|u-1/2|} \\ \iff & B \cdot \lambda \cdot \exp\left(\frac{-|F^{-1}_{\text{TLap}}(u)|}{\lambda}\right) - B \cdot \lambda \cdot \exp\left(\frac{-A}{\lambda}\right) = \underbrace{-|u - 1/2| + \frac{1}{2}}_{=\min\{u, 1-u\}} \end{aligned}$$

$$\begin{aligned}
&\iff \exp\left(\frac{-|F^{-1}_{\mathrm{TLap}}(u)|}{\lambda}\right) = \frac{\min\{u, 1-u\}}{B\cdot\lambda} + \exp\left(\frac{-A}{\lambda}\right) \\
&\iff |F^{-1}_{\mathrm{TLap}}(u)| = -\lambda\cdot\log\cdot\left[\frac{\min\{u, 1-u\}}{B\cdot\lambda} + \exp\left(\frac{-A}{\lambda}\right)\right] \\
&\iff F^{-1}_{\mathrm{TLap}}(u) = -\mathrm{sign}(u-0.5)\cdot\lambda\cdot\log\left[\frac{\min\{u, 1-u\}}{B\cdot\lambda} + \exp\left(\frac{-A}{\lambda}\right)\right].
\end{aligned}$$

We then sample $u \sim [0,1]$ uniformly at random, and compute $F^{-1}_{\mathrm{TLap}}(u)$ to obtain a sample from the Truncated Laplace distribution.

Details on the Numerical Experiments

To be added.