Privacy Accounting Pitfalls

Christian Janos Lebeda, Matthew Regehr, Gautam Kamath, and Thomas Steinke

University of Waterloo, Inria, Google Research, BARC

Privacy Accounting for ML

Privacy accounting: Compute differential privacy parameters ε and δ (Lower values implies stronger privacy).

For ML, the standard algorithm is DP-SGD. We achieve privacy by adding noise to gradients updates.

Algorithm: Differentially Private Stochastic Gradient Descent (DP-SGD)

Input : Dataset $(x_1, \dots, x_n)$.
Output: Privatized gradient updates: $\tilde{g}_1, \dots, \tilde{g}_T \in \mathbb{R}^d$

$(S_1, \dots, S_T) \leftarrow \mathcal{B}_{\gamma,T}(n)$ ▷ Sample T batches with sampling rate γ
for $t = 1, \dots, T$ **do**
  $\psi_t(\cdot) \leftarrow \mathcal{A}(g_1, \dots, g_{t-1}; \cdot)$ ▷ Update model and compute gradients
  $g_t \leftarrow \sum_{i \in S_t} \text{clip}(\psi_t(x_i))$ ▷ Clip and aggregate gradients
  $\tilde{g}_t \leftarrow g_t + e_t$ where $e_t \sim \mathcal{N}(0, \sigma^2 I_d)$ ▷ Add noise
return $(\tilde{g}_1, \dots, \tilde{g}_T)$

We point out technical details and common mistakes in privacy accounting of DP-SGD.

Poisson vs WOR Subsampling

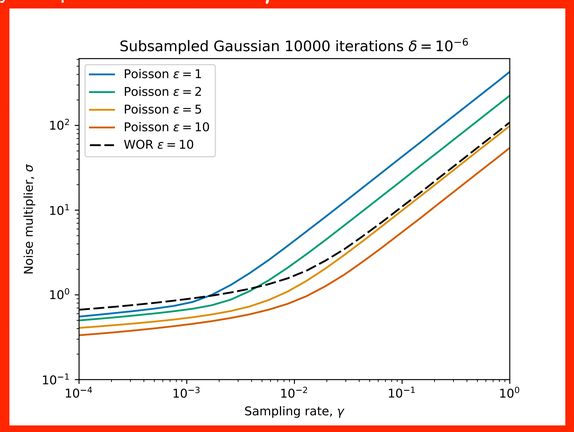
Takeaway: The choice of sampling scheme impacts differential privacy guarantees.

It is common to assume that batches are generated using Poisson subsampling. But this is often not the case in practice. We compared the privacy guarantees when using a very similar sampling scheme.

- Poisson subsampling takes a dataset and keeps each element independently w.p. γ.
- Without-replacement (WOR) subsampling uniformly samples a batch of size $\gamma \cdot n$.

δ	ε (Poisson)	ε (WOR)
10^{-7}	1.19	17.48
10^{-6}	0.96	15.26
10^{-5}	0.80	12.98
10^{-4}	0.64	10.62

Comparison of ε for Poisson and WOR with identical hyperparameters.

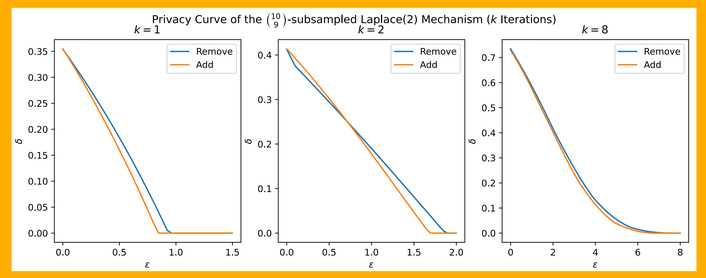


We show that WOR requires **twice** as much noise as Poisson subsampling.
For some hyperparameters the privacy guarantees differ **significantly** between the sampling schemes.

No WC Datasets under Composition

Takeaway: Privacy accounting is often unintuitive and we must use mathematical rigour.

Is it enough to find the worst-case pair of datasets for $k = 1$ iteration and all $\varepsilon \geq 0$?
Many implementations implicitly assumes this pair can be used for $k > 1$.



For Laplacian noise, we show this is not enough. For Gaussian noise, we conjecture it is sufficient.

Bigger Picture: Concurrent work

Takeaway: Any discrepancy between theory and practice should be avoided in differential privacy.

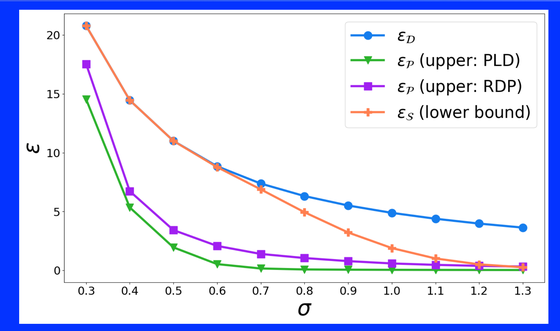
Concurrent independent work found large gaps in ε between Poisson subsampling and shuffling.

A recent paper proved that random allocation has similar privacy guarantees to Poisson.

Better algorithms are still needed for tight privacy accounting for these sampling schemes.

[CGKKMSZ24] How Private are DP-SGD Implementations?

[FS25] Privacy amplification by random allocation



More Details and Results in Paper

We give dominating pairs of datasets for the add and remove neighbouring relations.

We show that tight privacy accounting for WOR under the substitution (replace-one) neighboring relation is even more challenging.