Elliptical Perturbations for Differential Privacy

Matthew Reimherr and Jordan Awan
Department of Statistics, Pennsylvania State University, University Park, PA

Overview

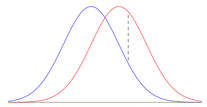
Objective: Study elliptical processes in locally convex vector spaces for differential privacy

Contributions:

- Condition for equivalence/orthogonality of elliptical distributions
- Criteria to determine whether an elliptical distribution satisfies ϵ-DP in any finite-dimensional space
- No elliptical distribution satisfies ϵ-DP in infinite dimensions
- Criteria for when elliptical distribution satisfies (ϵ, δ)-DP

Differential Privacy: [DMNS06]

- Let $\mathcal{X}$ be a set, $\mathcal{P} = \{P_{\underline{x}} \mid \underline{x} \in \mathcal{X}^n\}$ be a set of probability measures
- $\mathcal{P}$ satisfies **ϵ-Differential Privacy** (ϵ - DP) if for all measurable sets B and all $\underline{x}, \underline{x}'$ differing in one entry, we have $P_{\underline{x}}(B) \leqslant e^{\epsilon} P_{\underline{x}'}(B)$



Elliptical Distribution

A measure, P, over a (locally convex) vector space $\mathbb{X}$ is *elliptical* [Fan17] if its characteristic functional, $\widetilde{P} : \mathbb{X}^* \to \mathbb{C}$, has the form

$$\widetilde{P}(g) = \int_{\mathbb{X}} \exp\{ig(x)\}\, dP(x) = e^{ig(\mu)} \phi_0(C(g,g)) \qquad g \in \mathbb{X}^*,$$

- $\mathbb{X}^*$: dual space of $\mathbb{X}$
- $\mu \in \mathbb{X}$: mean or center of distribution
- $C : \mathbb{X}^* \times \mathbb{X}^* \to \mathbb{R}$: covariance or dispersion of distribution
- ϕ_0 : positive definite, continuous at 0, $\phi_0(0) = 1$

Connection to Gaussian Processes

If $X \in \mathbb{X}$ is elliptical then it can be written as [Fan17]

$$X \overset{\mathcal{L}}{=} \mu + VZ$$

- $V \in \mathbb{R}$ is random with distribution ψ
- $Z \in X$ is Gaussian with covariance C
- V and Z are independent
- $\mathcal{E}(\mu, C, \psi)$ denotes the distribution of X

Theorem (Equivalence of Measures)

Let $P_1 \sim \mathcal{E}(\mu_1, C, \psi)$ and $P_2 \sim \mathcal{E}(\mu_2, C, \psi)$ be two elliptical measures over a locally convex topological vector space, $\mathbb{X}$. P_1 and P_2 are equivalent if $\mu_1 - \mu_2$ lies in the Cameron-Martin space of C and orthogonal otherwise.

Significance

- Same condition as for Gaussian processes [Bog98, Theorem 2.4.5]
- For privacy, summary $T(D)$ must reside in Cameron-Martin space
- Privacy disclosure occurs with probability one otherwise

Theorem (Finite Dimensional Privacy)

Assume $\mathbb{X} = \mathbb{R}^d$ and that $\widetilde{T}_D$ has density proportional to

$$f_{\widetilde{T}_D}(x) \propto f(\sigma^{-2}(x - T_D)^\top \Sigma^{-1}(x - T_D)),$$

where $f : [0, \infty) \to [0, \infty]$ is a decreasing function. Set

$$\Delta = \sup_{D \sim D'} \|\Sigma^{-1/2}(T_D - T_{D'})\|_2.$$

If $\Delta < \infty$, $f(0) < \infty$, and $\limsup_{c \to \infty} \frac{f((c-\Delta)^2)}{f(c^2)} < \infty$,

then $\widetilde{T}_D$ satisfies ϵ-DP, where $\exp(\epsilon) = \sup_{c \geqslant \sigma^{-1}\Delta} \frac{f((c - \sigma^{-2}\Delta)^2)}{f(c^2)} < \infty.$

Theorem (Infinite Dimensional Privacy)

- *Let $T : \mathcal{D} \to \mathbb{X}$ and $\widetilde{T}_D = T_D + \sigma X$, where $X \sim \mathcal{E}(0, C, \psi)$. If $\mathbb{X}$ is infinite dimensional then $\widetilde{T}_D$ will not achieve ϵ-DP for any choice of σ.*
- *However, for any $\epsilon > 0$ and $\delta > 0$,*

$$\widetilde{T}_D = T_D + \sigma X, \qquad \text{with} \qquad \sigma^2 \geqslant \frac{2\log(2/\delta')}{\epsilon^2} \Delta^2$$

achieves (ϵ, δ)-DP, where δ' satisfies $\delta = 2M_V(\log(\delta'/2))$ and M_V is the moment generating function of mixing coefficient V.

Interpretation

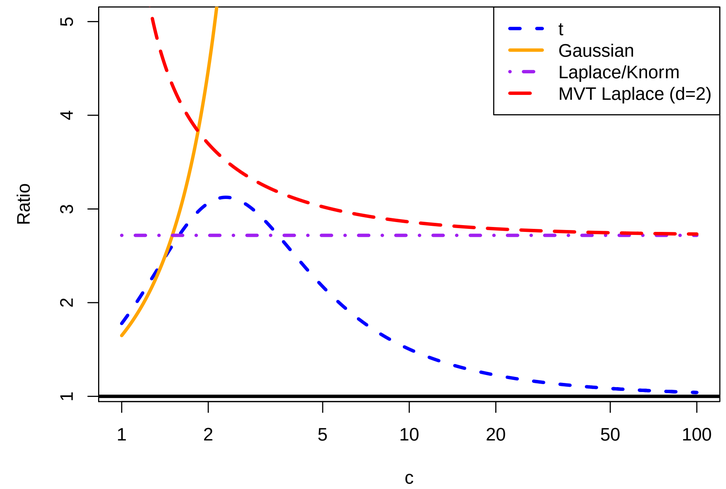
- In $\mathbb{R}^d$, one-dimensional criteria to determine ϵ-DP
- In infinite dimensions, the value V can be "inferred" from $\widetilde{T}_D$
- As far as privacy goes, all elliptical distributions behave like Gaussian in infinite dimensional spaces (see [MRS19, HRW13])

Examples

- For t, Gaussian, Laplace, and Multivariate Laplace, plot

$$\frac{f((c-\Delta)^2)}{f(c^2)}, \quad \text{for } c \in (1, 100), \text{ using } \Delta = 1.$$

- If the ratio is bounded, then it satisfies $\epsilon - DP$ for some ϵ
- For Gaussian, ratio goes to infinity as $c \to \infty$
- MVT Laplace has an asymptote at $c = \Delta$
- For Laplace, the ratio is constant
- For t, the ratio converges to 1



Discussion

- [BS19] show that alternative distributions can be useful for privacy in 1-dim. Could considering elliptical versions of these distributions
- Highlights the need to study privacy for complex data structures; multivariate tools do not always extend as expected

References

[Bog98] Vladimir Igorevich Bogachev. *Gaussian measures*. Number 62. American Mathematical Soc., 1998.

[BS19] Mark Bun and Thomas Steinke. Average-case averages: Private algorithms for smooth sensitivity and mean estimation. *arXiv preprint arXiv:1906.02830*, 2019.

[DMNS06] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. *Calibrating Noise to Sensitivity in Private Data Analysis*, pages 265–284. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[Fan17] Kai Wang Fang. *Symmetric Multivariate and Related Distributions: 0*. Chapman and Hall/CRC, 2017.

[HRW13] R. Hall, A. Rinaldo, and L. Wasserman. Differential privacy for functions and functional data. *The Journal of Machine Learning Research*, 14(1):703–727, 2013.

[MRS19] Ardalan Mirshani, Matthew Reimherr, and Aleksandra Slavković. Formal privacy for functional data with gaussian perturbations. In *International Conference on Machine Learning*, pages 4595–4604, 2019.

NSF Awards SES-1534433, SES-1853209, and DMS-1712826 to The Pennsylvania State University

mreimherr@psu.edu