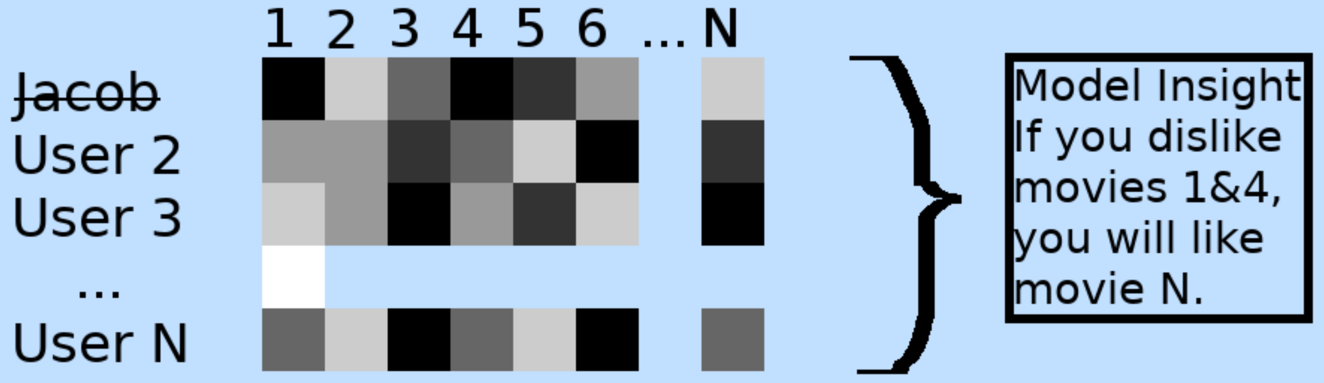
Capacity Bounded Differential Privacy

Kamalika Chaudhuri Jacob Imola Ashwin Machanavajjhala

Intoduction

Privacy risks in ML: Richer model space, more details about small populations in data

Example: Jacob partakes in study on movie DBs



Jacob's Public Blog:

Post 1:	Post 2:	Post 3:
Movie 1 had terrible acting!	Movie 5 deserves 2 thumbs up!	Movie 4 was less exciting than watching paint dry!

Adversary with blog and model insight could leak Jacob's Movie N rating!

Formalization: Let D,D' be two DBs one with population P and one without. Then, learner M satisfies (α, ε)-Renyi Differential Privacy (RDP) if:

$$R_\alpha(M(D), M(D')) \leq e^{(\alpha-1)\epsilon}$$

Interpretation: P cannot change the distribution of M(D), so no one can even tell if P is in D.

Capacity Bounded Differential Privacy

Renyi-DP can be written dually (for some f*) as

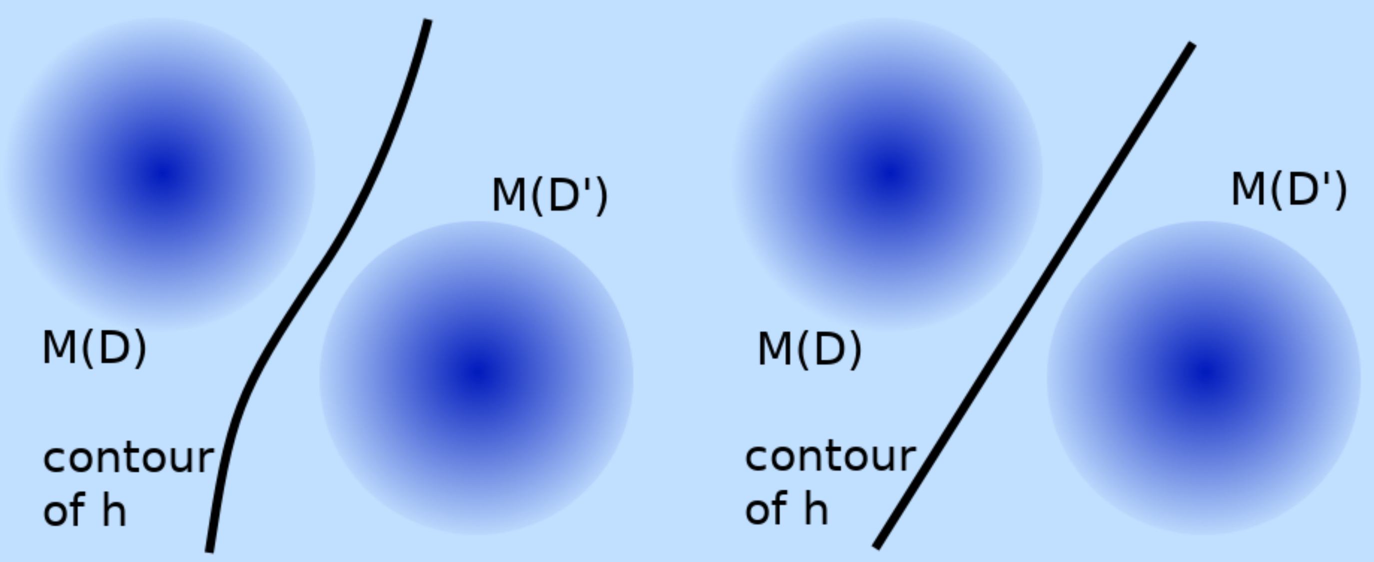
$$\sup_{h:\Omega\to\mathbb{R}} \left(\mathop{\mathbb{E}}_{x\sim M(D)}[h(x)] - \mathop{\mathbb{E}}_{x\sim M(D')}[f^*(h(x))] \right) \leq e^{(\alpha-1)\epsilon}$$

Same definition as *f*-GAN

Interpretation: How well can adversarial h distinguish M(D) and M(D')?

Key Idea: In certain cases we know a restriction on *h* which limits its power to distinguish

Complex, optimal *h* Simple, less powerful *h*

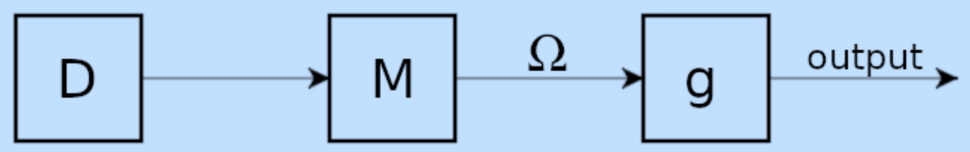


Definition (*H*-capacity bounded Renyi-DP):

$$\sup_{h\in\mathcal{H}} \left(\mathop{\mathbb{E}}_{x\sim M(D)}[h(x)] - \mathop{\mathbb{E}}_{x\sim M(D')}[f^*(h(x))] \right) \leq e^{(\alpha-1)\epsilon}$$

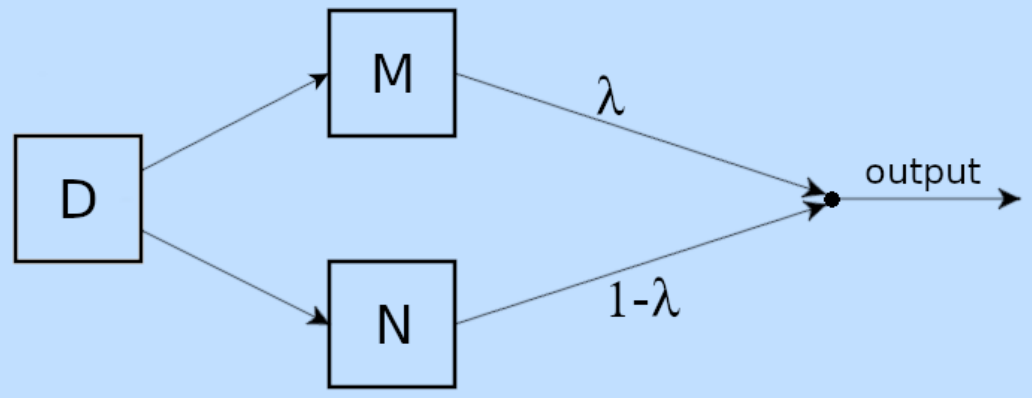
Properties of CBP

Let M,N satisfy *H*,*K*-CBP with params ε, γ respectively



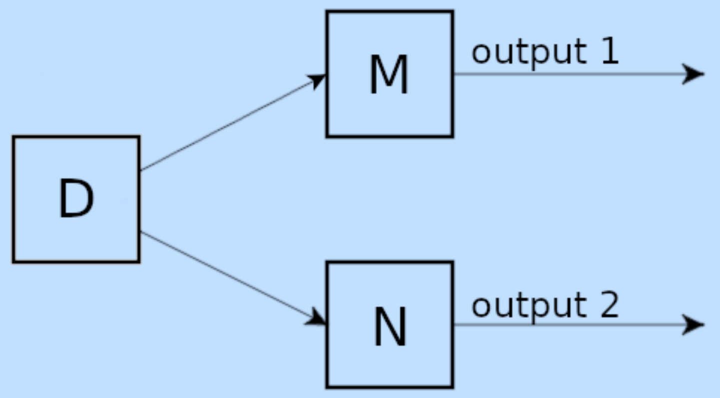
Post-processing: *H*,*G*,*I* function classes such that $I \circ G \subseteq H$. Then, g∘M satisfies *I*-CBP with param ε.

Standard DP makes no restriction on G but that is necessary here



Convexity: *H=K* and O is model that returns M w.p. λ and N w.p. 1-λ. Then, O satisfies *H*-CBP with parameter $\lambda\epsilon + (1-\lambda)\gamma$.

Same result as standard DP



Composition: *G* is $\{h+k \mid h\in H, k\in K\}$, O is the model (M,N). Then, O satisfies *G*-CBP with parameter ε + γ.

Standard DP allows adaptive composition (N can look at M output)

Mechanisms

Gaussian: For $D\in\mathbb{R}^d$:

M(D) = D + 

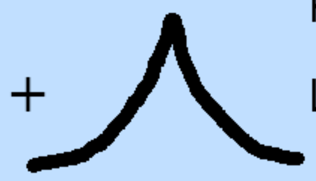
i.i.d. std = σ

Renyi DP: $\frac{\alpha}{2\sigma^2}$

Lin-CBP: at most

$$\frac{1}{\alpha-1}\log\left(1+\frac{2^{d(\alpha-1)}\sqrt{\pi/2}^{\alpha-1}}{\sigma^\alpha}\right)$$

Laplace: For $D\in\mathbb{R}^d$:

M(D) = D + 

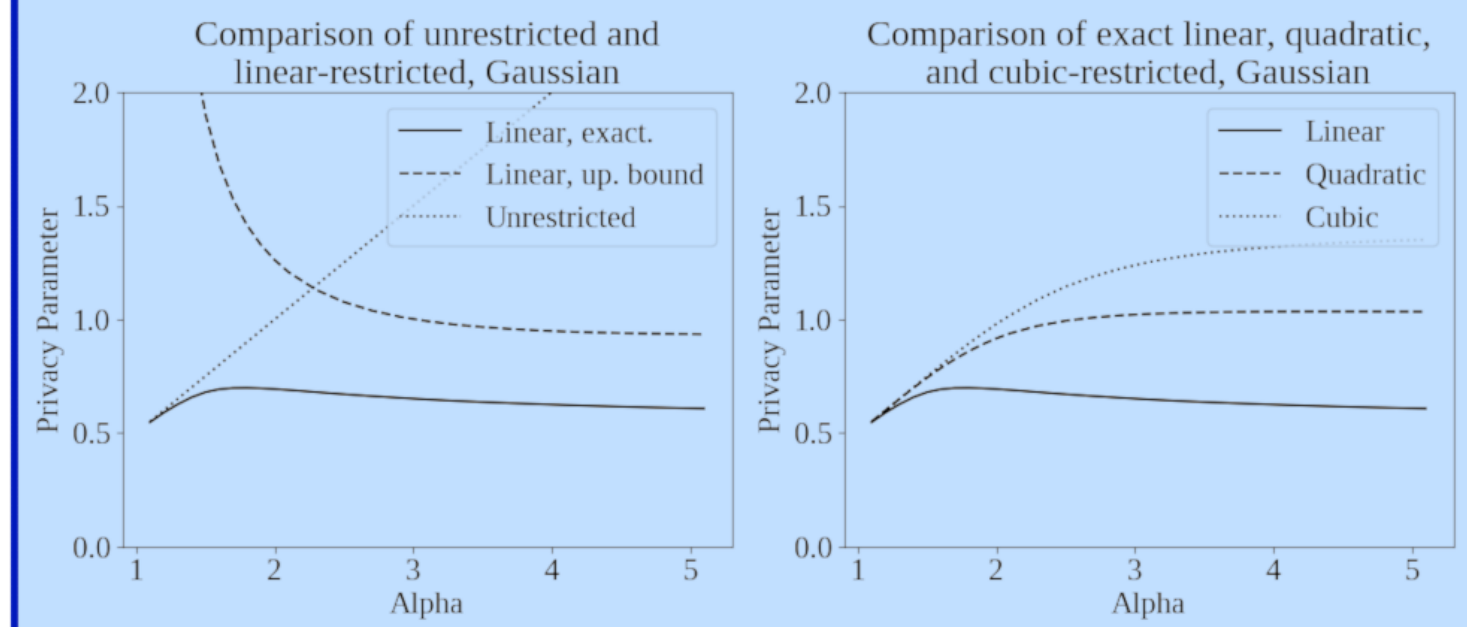
i.i.d. std = 1/ε

Renyi DP: at least $\epsilon - \frac{d\log(2)}{\alpha-1}$

Lin-CBP: at most

$$\frac{1}{\alpha-1}\log\left(1+2^{d(\alpha-1)}\epsilon^\alpha\right)$$

Mechanism Performance Plots



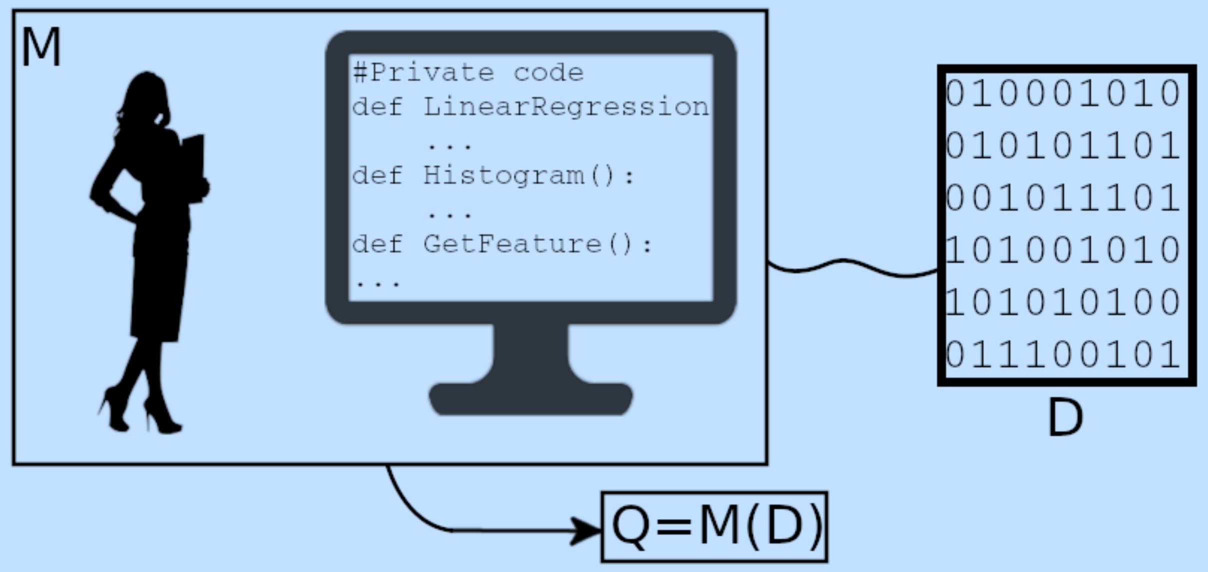
ε is much smaller for Poly-CBP

Subtle points: Poly-CBP eventually decreases with α

Our lin-CBP bound is worse than RDP for small α

Generalization Properties

Concrete Example: Analyst produces model Q using private library with access to D



Theorem: If M satisfies capacity-bounded DP then empirical loss and true loss are close:

$$\left| \mathop{\mathbb{E}}_{D\sim\mathcal{D}^d} \left(\frac{1}{d}\sum_{i=1}^{d} \ell(Q, x_i) - \mathop{\mathbb{E}}_{x\sim\mathcal{D}}[\ell(Q, x_i)] \right) \right| \leq 8\sqrt{\epsilon}$$

Conclusion

- -CBP is novel privacy approach when adversary is known to be bounded
- -Satisfies many information-theoretic properties that RDP satisfies
- -For simple adversaries can prove stronger privacy than RDP for the same algorithm

Future Steps

- -Find M which is not RDP but is H-CBP for finite parameter ε
- -Make composition result adaptive
- -Determine ε for more complex H
- -Prove other generalization properties