

DIFFERENTIAL PRIVACY: GENERAL INFERRENTIAL LIMITS VIA INTERVALS OF MEASURES

James Bailie¹, Ruobin Gong²

¹jamesbailie@g.harvard.edu – Harvard University, USA ²ruobin.gong@rutgers.edu – Rutgers University, USA

DIFFERENTIAL PRIVACY AS LIPSCHITZ CTY.

Let $M : \mathcal{X} \times [0, 1] \rightarrow \mathcal{T}$ be a data-release mechanism with each dataset $x \in \mathcal{X}$ inducing a probability P_x on \mathcal{T} .

Definition. (Dwork et al. 2006) Given a data universe \mathcal{X} equipped with a metric d , the mechanism M satisfies ϵ -differential privacy (DP) if

$$d_{\text{MULT}}(P_x, P_{x'}) \leq \epsilon d(x, x'),$$

for all $x, x' \in \mathcal{X}$, where

1. $d_{\text{MULT}}(P, Q) = \sup_S \left| \ln \frac{P(S)}{Q(S)} \right|$ is the *multiplicative distance* between measures P, Q on \mathcal{T} ;

2. $d(x, x')$ is the shortest path length between x and x' in a graph on \mathcal{X} with unit-length edges; for example:

- (bounded case) the *Hamming distance*

$$d_{\text{HAM}}(x, x') = \sum_{i=1}^n 1_{x_i \neq x'_i},$$

if $|x| = |x'| = n$, and ∞ otherwise, where the data $x = (x_1, x_2, \dots, x_n)$ are vectors and $|x|$ is the size of x ; or

- (unbounded case) the *symmetric difference* metric

$$d_{\Delta}(x, x') = |x \setminus x'| + |x' \setminus x|,$$

where the data $x, x' \in \mathcal{X}$ are multisets and $x \setminus x'$ is the (multi-)set difference.

EXAMPLES

1. **Randomised Response** (Warner 1965): Taking $\mathcal{X} = \bigcup_{n \in \mathbb{N}} \{0, 1\}^n$ as the data universe, and $d = d_{\text{HAM}}$, define the randomised response mechanism:

$$M_{\text{RR}}(x, U) = (\dots, x_i + U_i \bmod 2, \dots)$$

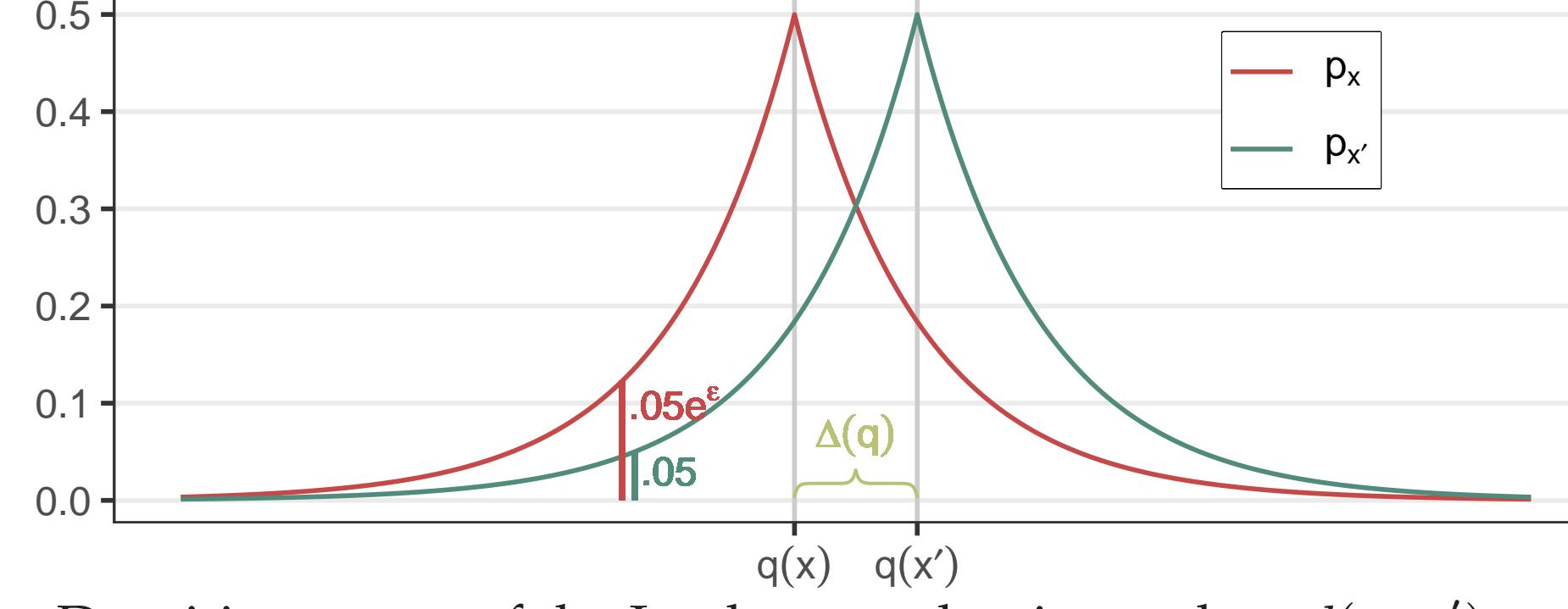
where $U_1, U_2, \dots \stackrel{iid}{\sim} \text{Bernoulli}(p)$. That is, given a binary n -vector x as input, M_{RR} outputs another binary n -vector with i -th component $x_i + U_i \bmod 2$, flipping each bit x_i with probability $p = (\exp \epsilon + 1)^{-1}$.

2. **The Laplace Mechanism** M_{Lap} adds noise to a query $q : \mathcal{X} \rightarrow \mathbb{R}^k$ with standard deviation proportional to its *global ℓ_1 -sensitivity* $\Delta(q)$, i.e.:

$$M_{\text{Lap}}(x, U) = q(x) + bU,$$

where $b = \Delta(q)/\epsilon$, and U is a k -vector of iid Laplace random variables with density $f(z) = 0.5 \exp(-|z|)$, and

$$\Delta(q) = \sup_{d(x, x')=1} \|q(x) - q(x')\|_1.$$



Densities $p_x, p_{x'}$ of the Laplace mechanism, when $d(x, x') = 1$

BOUNDS ON THE PRIVATISED DATA PROBABILITY

The relevant vehicle for inference in the private setting is the marginal probability of the observed data t (the **privatised data probability**):

$$P(t \in S | \theta) = \int_{\mathcal{X}} P_x(S) dP_{\theta}(x).$$

- Viewed as a function of θ , this is the *marginal likelihood* of θ .
- All frequentist procedures compliant with likelihood theory and all Bayesian inference from privatised data hinge on this function.

Theorem. Let M be ϵ -DP. If $\text{supp}(x | t, \theta)$ is d -connected, then for any $x_* \in \text{supp}(x | t, \theta)$,

$$p(t | \theta) \in p_{x_*}(t) \exp(\pm \epsilon d_*),$$

where $d_* = \sup_{x \in \text{supp}(x | t, \theta)} d(x, x_*)$. Furthermore if $\text{supp}(x | t, \theta)$ is d -connected for $P(t | \theta)$ -almost all $t \in \mathcal{T}$, then

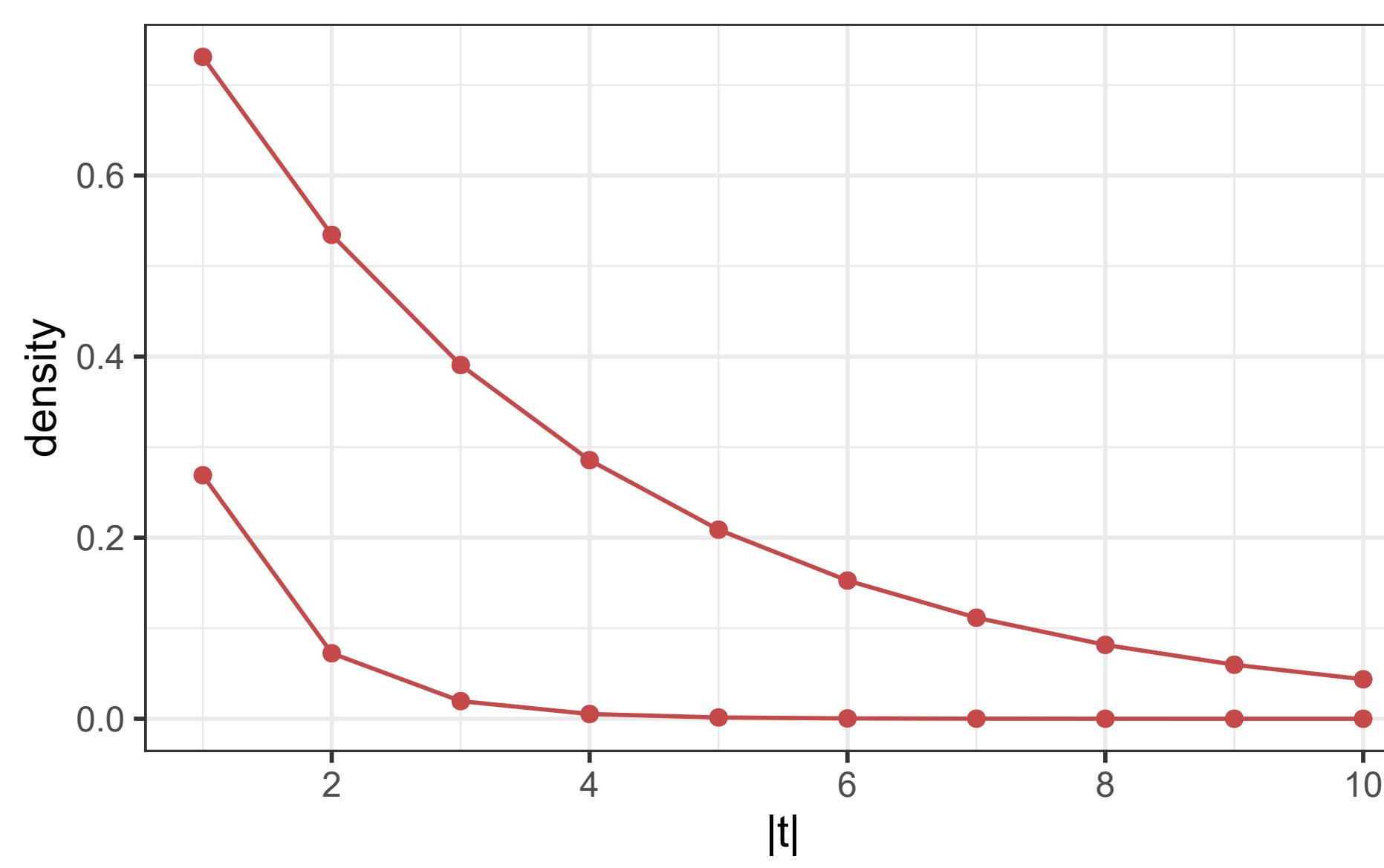
$$P(t | \theta) \in \mathcal{I}(L_{\epsilon}, U_{\epsilon}),$$

where L_{ϵ} and U_{ϵ} have densities

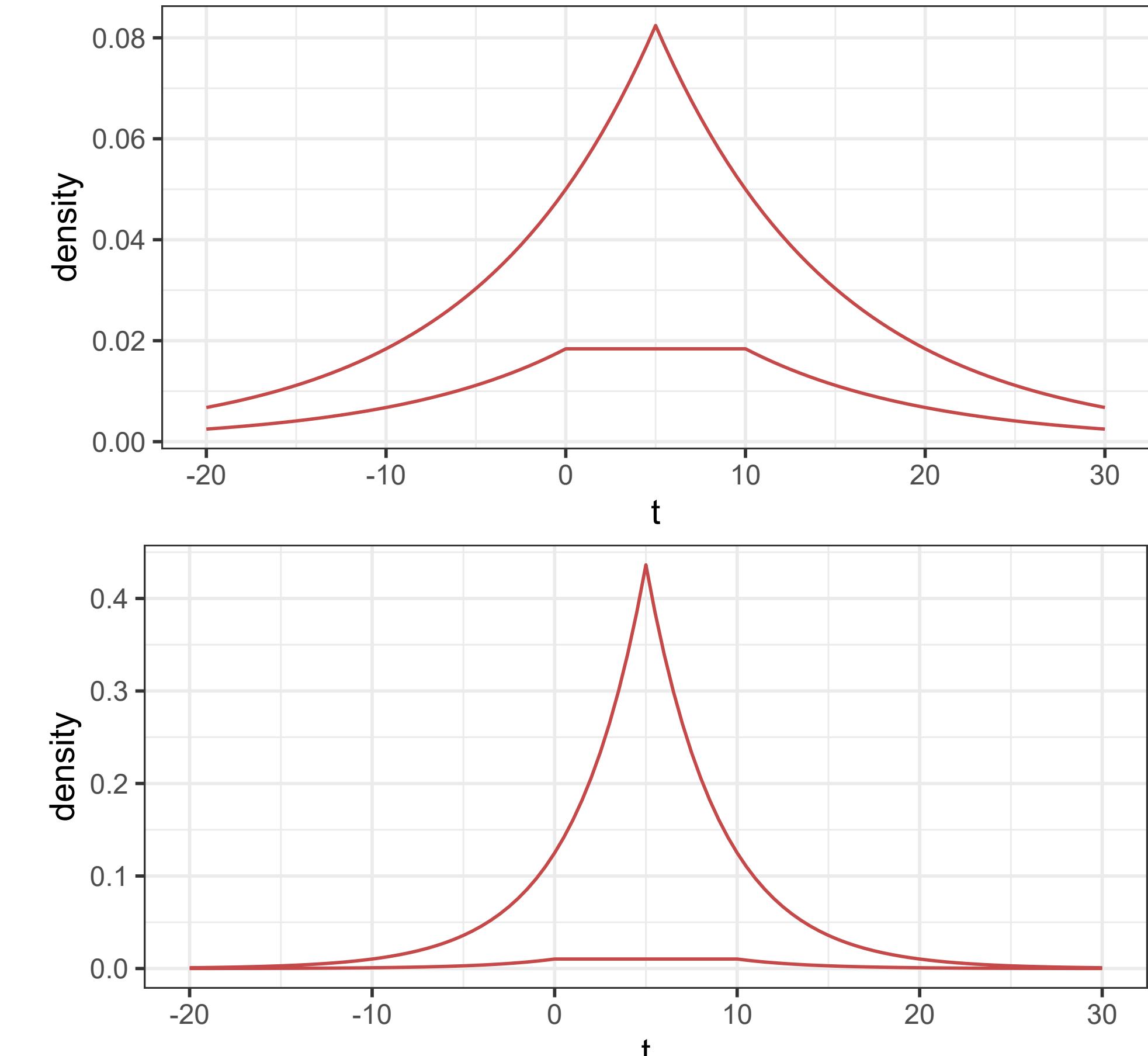
$$\text{ess sup}_{x_* \in \text{supp}(x | t, \theta)} \exp(-\epsilon d_*) p_{x_*} \text{ and } \text{ess inf}_{x_* \in \text{supp}(x | t, \theta)} \exp(\epsilon d_*) p_{x_*}.$$

Note that $\mathcal{I}(L_{\epsilon}, U_{\epsilon})$:

- depends on the data generating distribution P_{θ} only through $\text{supp}(x | t, \theta)$. When $\text{supp}(P_{\theta})$ is constant, it is completely *free of θ* .
- is *non-vacuous* whenever $d_* < \infty$. (For example, when the analyst has partial prior knowledge of the data X so that $|x| < \infty$ for all $x \in \text{supp}(P_{\theta})$.)



Example 1 (randomised response) illustrated. Upper and lower density bounds for the privatised data probability $p(t | \theta)$ with $\epsilon = 1$ and $\text{supp}(x | t, \theta) \subset \{x : |x| \leq 10\}$. These bounds are a function of t only through $|t|$ (the number of records).



Example 2 (the Laplace mechanism for a privatised binary sum) illustrated. Upper and lower density bounds for $p(t | \theta)$ with $\epsilon = 0.1$ (top) and $\epsilon = 0.25$ (bottom). Note that these bounds:

- do not depend on θ nor the assumed data model P_{θ} .
- are tighter and more informative when privacy protection is more stringent (smaller ϵ).

FREQUENTIST PRIVACY-PROTECTED INFERENCE

Theorem (Neyman-Pearson hypothesis testing). Consider testing

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta = \theta_1,$$

for some $\theta_0 \neq \theta_1 \in \Theta$. Let $S_i = \text{supp}(x | t, \theta_i)$ and suppose that every $x \in S_0$ is d -connected to every $x' \in S_1$.

In the private setting, the power of any level- α test is bounded above by

$$\alpha \exp(d_{**}\epsilon),$$

where $d_{**} = \sup_{x \in S_0, x' \in S_1} d(x, x')$.

This Theorem generalises the classical result of Wasserman and Zhou 2010 beyond the case of iid records.

BAYESIAN PRIVACY-PROTECTED INFERENCE

Suppose that $\text{supp}(x | t) := \bigcup_{\theta \in \text{supp}(\pi)} \text{supp}(x | t, \theta)$ is d -connected for $P(t)$ -almost all $t \in \mathcal{T}$. Also assume the prior π on θ is proper.

Theorem (prior predictive bounds). The analyst's prior predictive probability for $t \sim M(X, U)$ (that is ϵ -DP) satisfies

$$\underline{p}_{\epsilon}(t) \leq p(t) \leq \bar{p}_{\epsilon}(t),$$

for every $t \in \mathcal{T}$, where \underline{p}_{ϵ} and \bar{p}_{ϵ} are defined as

$$\text{ess sup}_{x_* \in \text{supp}(x | t)} \exp(-\epsilon d_*) p_{x_*} \text{ and } \text{ess inf}_{x_* \in \text{supp}(x | t)} \exp(\epsilon d_*) p_{x_*}$$

respectively, with $d_* = \sup_{x \in \text{supp}(x | t)} d(x, x_*)$.

Theorem (posterior bounds). The analyst's posterior probability given (a realisation of an ϵ -DP mechanism) t satisfies

$$\pi(\theta | t) \in \pi(\theta) \exp(\pm \epsilon d_{**}),$$

where $d_{**} = \sup_{x, x' \in \text{supp}(x | t)} d(x, x')$.

This Theorem elucidates ϵ -DP's guarantee of *prior-to-posterior* privacy (*restricting an attacker's posterior departure from their prior*, Duncan and Lambert 1986), under:

- arbitrary specifications of the data model P_{θ} ;
- arbitrary choice of (proper) prior $\pi(\theta)$; and
- is non-vacuous so long as d_{**} is finite (which is not unreasonable in general).

SUMMARY

- We provide general limits on important statistical quantities in *likelihood*, *frequentist* and *Bayesian* inference from ϵ -differentially private data.
- Under very mild assumptions, these results are valid for arbitrary ϵ -DP mechanisms M , parameters $\theta \in \Theta$, priors π and data generating models $P_{\theta}(x)$.
- Our bounds are *optimal* – they cannot be further improved without assumptions on M, θ, π or $P_{\theta}(x)$.
- Therefore, these bounds are useful representations of the limits of statistical learning – for attackers as well as valid analysts – under the constraint of ϵ -DP.
- These results were accomplished by characterising ϵ -DP using a foundational tool from the IP literature – the *interval of measures*.
- This work provides clarity to the *semantic debate on privacy and disclosure* in the curation and governance of official statistics.