Differentially Private Learning of Exponential Distributions: Adaptive Algorithms and Tight Bounds

Bar Mahpud
Bar-Ilan University

Or Sheffet
Bar-Ilan University

Abstract

We study the problem of learning exponential distributions under differential privacy. Given n i.i.d. samples from $\mathrm{Exp}(\lambda)$, the goal is to privately estimate λ so that the learned distribution is close in total variation distance to the truth. We present two complementary pure DP algorithms: one adapts the classical maximum likelihood estimator via clipping and Laplace noise, while the other leverages the fact that the $(1-1/e)$-quantile equals $1/\lambda$. Each method excels in a different regime, and we combine them into an adaptive best-of-both algorithm achieving near-optimal sample complexity for all λ. We further extend our approach to Pareto distributions via a logarithmic reduction, prove nearly matching lower bounds using packing and group privacy Karwa and Vadhan (2017), and show how approximate (ϵ, δ)-DP removes the need for externally supplied bounds. Together, these results give the first tight characterization of exponential distribution learning under DP and illustrate the power of adaptive strategies for heavy-tailed laws.

1 Introduction

The exponential distribution occupies a central role in probability and statistics. It arises naturally in modeling interarrival times of Poisson processes (Ross, 1996; Asmussen, 1987), waiting times in queueing systems (Kleinrock, 1975; Daskin, 2021), lifetimes in reliability analysis and survival studies (Smith, 2012; Johnston, 2012), and interarrival processes in communication and computer networks (Kleinrock, 1975; Bertsekas and Gallager, 1992). Its simplicity—being defined by a single rate parameter λ—makes it a canonical distribution both for theory and for practice. Moreover, the exponential law serves as a fundamental building block for more complex families, including the Pareto and Weibull distributions, which are widely used in economics, finance (Mandelbrot, 1960; Gabaix, 2016), and network modeling (Crovella and Bestavros, 1997; Feldmann et al., 1999). Accurate estimation of the rate parameter λ is therefore a fundamental statistical task.

In modern applications, however, the data from which such parameters must be estimated are often sensitive. Waiting times may reveal information about individual users, interarrival processes may correspond to private communication events, and survival data may capture confidential medical information. As a result, there is a pressing need for algorithms that allow reliable learning of the exponential distribution while protecting the privacy of individuals. *Differential privacy* (DP) (Dwork et al., 2006) has emerged as the gold standard for such guarantees, providing provable bounds on the information leakage about any single data point. Yet, despite significant progress on private distribution learning in general, the case of exponential distributions has not been studied in detail.

Our work lies at the intersection of private distribution learning and heavy-tailed statistics. Classical results on distribution learning under differential privacy include private mean and variance estimation, histograms, and quantiles (see, e.g., Dwork et al. (2006); Smith (2011); Bun and Steinke (2016); Bun et al. (2014)). More recent works have studied private parameter estimation for specific distribution families, such as Gaussian mean and covariance (Kamath et al., 2019, 2021), multinomials (Kifer et al., 2012; Acharya et al., 2017), and heavy-tailed distributions (Kamath et al., 2021; Bun et al., 2015). In particular, Karwa and Vadhan (2017) introduced finite sample lower bounds for private distribution learning via group privacy, which we adapt in our lower bound arguments. On the statistical side, exponential and Pareto laws have been extensively analyzed due to their role in reliability, survival analysis, and modeling of heavy tails. To the best of our knowledge, however, no prior work has provided differentially private algorithms with tight sample complexity guarantees for exponential or Pareto distributions. Our results thus fill this gap yield near-optimal private

learners in this fundamental setting.

In this work our goal is to design algorithms that, given n i.i.d. samples from $\mathrm{Exp}(\lambda)$, produce an estimate $\tilde{\lambda}$ such that the learned distribution $\mathrm{Exp}(\tilde{\lambda})$ is α-close in total variation distance to the true distribution in total variation distance while adhering to differential privacy. As in exhibit in Section 2.3, this requirement reduces to obtaining a multiplicative $(1 \pm \alpha)$-approximation to λ. The central question we address is therefore: how many samples are required to achieve this goal under differential privacy?

Towards that end, we develop in Section 3 two complementary algorithmic strategies. The first builds on the classical maximum likelihood estimator (MLE), adapting it to the private setting by clipping and perturbing the sample mean. The second leverages a unique structural property of the exponential distribution: its $(1 - 1/e)$-quantile is $1/\lambda$. By privately estimating this quantile, we obtain a direct estimator for λ without relying on the mean. Each approach dominates in a different regime of the parameter space: the MLE-based method performs better when λ is large, while the quantile-based method is superior when λ is small. We therefore propose a unified adaptive "best-of-both" ϵ-DP algorithm that combines their strengths. Then, in Section 4 we further extend our method to the Pareto distribution, a prototypical heavy-tailed distribution.

Our ϵ-DP algorithm requires an apriori knowledge of lower- and upper-bounds, $\lambda_{\min}$ and $\lambda_{\max}$ resp., on the value of λ. In section 5 we establish a nearly matching lower bound on our sample complexity showing that the doubly logarithmic $\log\log(\lambda_{\max}/\lambda_{\min})$ dependency is necessary, using group privacy arguments from Karwa and Vadhan (2017). Our lower-bound shows that our sample complexity are essentially tight, up to logarithmic factors.

Finally, in Section 6 we address the challenge of initialization. Bypassing the lower-bound, we give an approximate (ϵ, δ)-DP algorithm that retrieves such bounds on λ using the histogram-based algorithm of Vadhan (2017), thus removing the dependency on these apriori bounds on λ.

In summary, this work makes the following contributions:

- We introduce the first differentially private algorithms for learning exponential distributions, achieving $(1 \pm \alpha)$-multiplicative accuracy of the rate parameter.
- We extend our techniques to the Pareto distribution, demonstrating the transferability of exponential learning to heavy-tailed distributions.
- We establish lower bounds showing that our sample complexities are tight up to logarithmic factors, and require apriori knowledge of bounds on the λ-parameter.
- We show how approximate-DP enables the removal of these apriori bounds by privately estimating initialization bounds.

Taken together, these results highlight the exponential distribution as a rich and instructive case study in private statistics. Our findings not only resolve its sample complexity up to logarithmic terms but also provide new algorithmic tools and structural insights that can and should extend to broader classes of distributions.

2 Preliminaries

2.1 Differential Privacy

Definition 2.1 (Differential Privacy)**.** A randomized algorithm $\mathcal{M} : \mathcal{X}^n \to \mathcal{Y}$ is (ϵ, δ)-*differentially private* (approximate differentially private) if for all neighboring datasets $D, D' \in \mathcal{X}^n$ and all measurable $S \subseteq \mathcal{Y}$,

$$\Pr[\mathcal{M}(D) \in S] \;\le\; e^{\epsilon} \Pr[\mathcal{M}(D') \in S] + \delta$$

when $\delta = 0$ we say that the algorithm is ϵ-*differentially private* (pure differentially private).

Definition 2.2 (Laplace Mechanism)**.** Let $f : \mathcal{X}^n \to \mathbb{R}^d$ be a function with ℓ_1-sensitivity

$$\Delta f \;=\; \max_{D, D'} \|f(D) - f(D')\|_1$$

over neighboring datasets D, D'. The *Laplace mechanism* releases

$$\mathcal{M}(D) \;=\; f(D) + (Z_1, \ldots, Z_d), \qquad Z_i \overset{\text{i.i.d.}}{\sim} \mathrm{Lap}(\Delta f/\epsilon)$$

This mechanism is ϵ-differentially private.

Lemma 2.3 (Basic Composition)**.** If $\mathcal{M}_1$ is (ϵ_1, δ_1)-DP and $\mathcal{M}_2$ is (ϵ_2, δ_2)-DP, then their sequential composition $(\mathcal{M}_1, \mathcal{M}_2)$ is $(\epsilon_1 + \epsilon_2,\ \delta_1 + \delta_2)$-DP. In particular, composing k pure ϵ-DP mechanisms yields $k\epsilon$-DP.

2.2 Probabilistic Background

Definition 2.4 (Total Variation Distance)**.** For two probability distributions P and Q over a measurable space $(\mathcal{X}, \mathcal{F})$, the *total variation distance* is defined as

$$\mathrm{TV}(P, Q) \;=\; \sup_{A \in \mathcal{F}} |P(A) - Q(A)| = \tfrac{1}{2} \int_{\mathcal{X}} |p(x) - q(x)|\, dx$$

Lemma 2.5 (Dvoretzky–Kiefer–Wolfowitz (DKW) Inequality)**.** Let $X_1, \ldots, X_n \sim P$ i.i.d. with CDF F, and let F_n denote the empirical CDF. Then for any $\eta > 0$,

$$\Pr\left[\sup_x |F_n(x) - F(x)| > \eta\right] \;\le\; 2e^{-2n\eta^2}$$

Lemma 2.6 (Multiplicative Chernoff Bound)**.** Let $X_1, \ldots, X_n$ be independent $\{0,1\}$-valued random variables with mean $\mu = \mathbb{E}[\sum_i X_i]$. Then for any $\alpha \in (0,1)$,

$$\Pr\left[\sum_i X_i \geq (1+\alpha)\mu\right] \leq \exp(-\tfrac{\alpha^2}{3}\mu)$$

$$\Pr\left[\sum_i X_i \leq (1-\alpha)\mu\right] \leq \exp(-\tfrac{\alpha^2}{2}\mu)$$

2.3 Exponential Distribution

For $\lambda > 0$, the exponential distribution $\mathrm{Exp}(\lambda)$ has density $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$. Its mean is $1/\lambda$, and its CDF is $F(x) = 1 - e^{-\lambda x}$.

Definition 2.7 (Maximum Likelihood Estimator for the Exponential Distribution)**.** Let $X_1, \ldots, X_n \sim \mathrm{Exp}(\lambda)$ be i.i.d. samples. The *maximum likelihood estimator (MLE)* for the rate parameter λ is

$$\hat{\lambda} \;=\; \frac{1}{\bar{X}} \;=\; \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^{-1}$$

We also recall below a fact and a corollary relating the exponential family and total variation, which will be used throughout; the proofs of both statements are deferred to the appendix for completeness (See Theorem A.1 and Theorem A.2).

Fact 2.8. Let $\mathrm{Exp}(\lambda_1), \mathrm{Exp}(\lambda_2)$ be two exponential distributions and assume $\lambda_1 \geq \lambda_2$. Denote $a = \frac{\ln(\lambda_1) - \ln(\lambda_2)}{\lambda_1 - \lambda_2}$, then:

$$TV(\mathrm{Exp}(\lambda_1), \mathrm{Exp}(\lambda_2)) = e^{-\lambda_2 \cdot a} - e^{-\lambda_1 \cdot a}$$

Corollary 2.9. If the rate parameters satisfy $\lambda_1 \in (1 \pm \alpha)\lambda_2$, then

$$TV(\mathrm{Exp}(\lambda_1), \mathrm{Exp}(\lambda_2)) \leq \alpha$$

2.4 Pareto Distribution

The *Pareto distribution* is a classical heavy-tailed law that is parameterized by a minimum (scale) parameter $x_m > 0$ and a shape parameter $\alpha_p > 0$, and is supported on $[x_m, \infty)$. Its probability density function is

$$f(x) \;=\; \alpha_p \cdot \frac{x_m^{\alpha_p}}{x^{\alpha_p+1}} \qquad \text{for } x \geq x_m.$$

We denote this law by $\mathrm{Pareto}(x_m, \alpha_p)$.

Remark. If $X \sim \mathrm{Pareto}(x_m, \alpha_p)$, then

$$Y \;=\; \ln\!\left(\tfrac{X}{x_m}\right) \;\sim\; \mathrm{Exp}(\alpha_p).$$

Thus, estimation of α_p can be reduced to exponential learning.

3 Pure Differentially Private Algorithms for Exponential Distribution Learning

We now turn to our main technical contribution: learning the parameter λ of an exponential distribution under pure ϵ-differential privacy. Our target is to output an estimate $\tilde{\lambda}$ such that the learned distribution $\mathrm{Exp}(\tilde{\lambda})$ is close in total variation distance to the true distribution $\mathrm{Exp}(\lambda)$. By Corollary 2.9, it suffices to achieve a $(1 \pm \alpha)$–multiplicative approximation of λ, since this immediately implies $\mathrm{TV}(\mathrm{Exp}(\tilde{\lambda}), \mathrm{Exp}(\lambda)) \leq \alpha$.

We develop two distinct strategies. The first method, described in Section 3.1, builds on the classical maximum–likelihood estimator (MLE), combined with careful clipping and noise addition. The second method, presented in Section 3.2, exploits a structural property of the exponential law: the $(1 - 1/e)$-quantile equals exactly $1/\lambda$, so privately estimating this quantile yields λ directly. The MLE-based pipeline is statistically powerful for larger values of λ, while the quantile-based estimator shines when λ is small. This motivates a unified *best-of-both algorithm* (Section 3.3) that adaptively chooses the better option.

3.1 Learning λ from the MLE

Our first subroutine, Algorithm 1, provides a way to approximate distributional quantiles in a private manner, which will serve as a building block for both range estimation and later for parameter learning. The idea is to scan over dyadic intervals, perturb empirical CDF values with Laplace noise, and return the smallest point where the noisy empirical CDF exceeds a perturbed threshold.

Algorithm 1 Unbounded Quantile Estimation

Input: $P = \{x_i\}_{i=1}^n \sim Exp(\lambda)$, bounds $0 < \lambda_{\min} < \lambda_{\max}$, quantile parameter $\theta \in (0,1)$, privacy budget $\epsilon > 0$.

procedure QUANTILE-ESTIMATION$(P, \lambda_{\min}, \lambda_{\max}, \theta, \epsilon)$

  $\tilde{T} \leftarrow (1-\theta) + Lap(\frac{2}{\epsilon n})$

  for $i = 0, 1, 2 \ldots$ **do**

    if $\frac{1}{n}|\{x_i \in P \mid \; x_i < 2^i \cdot \lambda_{\min}\}| + Lap(\frac{4}{\epsilon n}) \geq \tilde{T}$ **then**

      return $2^i \cdot \lambda_{\min}$

  return $\perp$

Lemma 3.1 (Privacy of Algorithm 1)**.** Algorithm 1 is ϵ-differentially private.

The proof is deferred to Theorem A.3.

Lemma 3.2 (Utility of Algorithm 1)**.** Algorithm 1 returns a 6-approximation of the $(1-\theta)$-quantile of P

w.p. $\geq 1-\beta$, given that $1/10 \leq \theta \leq 9/10$ and:

$$n \geq \max\left\{\frac{5}{\epsilon}\ln\left(\frac{4\log\left(\lambda_{\max}/\lambda_{\min}\right)}{\beta}\right), 200\ln\left(4/\beta\right)\right\}$$

Proof sketch. The analysis combines two ingredients: (i) the Dvoretzky–Kiefer–Wolfowitz inequality, which bounds the deviation of the empirical CDF from the true CDF; and (ii) concentration of Laplace noise, which controls the perturbations added to the CDF evaluations and the threshold. By ensuring both sampling error and noise error are at most $1/20$ (with probability at least $1-\beta$), we show that the returned quantile is within a constant factor of the true $(1-\theta)$-quantile. A detailed proof is provided in Theorem A.4. □

Corollary 3.3. Let $\tilde{Q}_{1-\theta}$ be the output of Algorithm 1. Fix a target failure probability $2\beta \in (0,1)$. There exists an absolute constant $c_0 \geq 1$ from Theorem 3.2 such that, if we set

$$\tilde{R} \;=\; C\,\tilde{Q}_{1-\theta}\ln n \quad \text{with} \quad C \;\geq\; \frac{c_0}{\ln(1/\theta)}\Big(1+\frac{\ln(1/\beta)}{\ln n}\Big)$$

then with probability at least $1-2\beta$ all n samples lie in $[0,\tilde{R}]$. In particular,

$$\tilde{R} \;=\; O\Big(\frac{1}{\lambda}\ln\Big(1/\theta\Big)\,\ln n\Big)$$

The proof is deferred to Theorem A.5.

Having obtained $\tilde{R}$, a private estimate of the range of the datapoints, we next turn to estimating λ via the maximum likelihood approach. Since the exponential MLE is simply the inverse of the sample mean, our strategy is to clip the data to the estimated range and then release a noisy mean under the Laplace mechanism. This yields Algorithm 2.

Algorithm 2 Private Exponential Distribution MLE

Input: $P = \{x_i\}_{i=1}^n \sim Exp(\lambda)$, range R, privacy budget $\epsilon > 0$.
procedure PRIVATE-MLE(P, R, ϵ)
  $\forall i:\ \bar{x}_i \leftarrow \min\{x_i, R\}$ ▷ Clipping
  $\tilde{s} \leftarrow \frac{1}{n}\sum_{i=1}^{n}\bar{x}_i + Lap(\frac{R}{\epsilon n})$
  $\tilde{\lambda} \leftarrow \frac{1}{\tilde{s}}$
  return $\tilde{\lambda}$

Lemma 3.4 (Privacy of Algorithm 2)**.** Algorithm 2 is ϵ-differentially private.

The proof is deferred to Theorem A.6.

Lemma 3.5 (Utility of Algorithm 2)**.** Algorithm 2 returns an $(1\pm\alpha)$-approximation of λ w.p. $\geq 1-\beta$, given that:

$$n \geq O\left(\max\left\{\frac{R\ln(\frac{1}{\beta})}{\epsilon(\alpha-e^{-\lambda R})}, \frac{\ln(\frac{1}{\beta})}{\alpha^2}\right\}\right)$$

Proof sketch. The estimator $\tilde{\lambda}$ is defined as the reciprocal of the privatized clipped mean. Its error arises from three sources: (i) sampling error of the empirical mean, controlled by a multiplicative Chernoff bound for exponentials; (ii) clipping bias, which is at most $e^{-\lambda R}/\lambda$ since the exponential tail mass beyond R is $\exp(-\lambda R)$; and (iii) Laplace noise added for privacy, which with probability $1-\frac{\beta}{2}$ is bounded by $\frac{R}{\epsilon n}\ln(\frac{2}{\beta})$. Combining these bounds shows that $\tilde{s}$, the noisy clipped mean, is within an additive α/λ of the true mean provided

$$n \;\geq\; O\Big(\max\Big\{\tfrac{R\ln(1/\beta)}{\epsilon(\alpha-e^{-\lambda R})},\ \tfrac{\ln(1/\beta)}{\alpha^2}\Big\}\Big)$$

under the assumption[1] $\alpha > e^{-\lambda R}$. Taking reciprocals then yields that $\tilde{\lambda}$ is an $(1\pm\alpha)$-approximation to λ with probability at least $1-\beta$. Full details are deferred to Theorem A.7. □

We now combine the two ingredients: the quantile-based range estimator and the private MLE. The resulting end-to-end algorithm takes raw samples, first derives a safe clipping range, and then computes a private estimate of λ using the clipped noisy mean. The complete pipeline is summarized in Algorithm 3 below.

Algorithm 3 Private Exponential Learning via MLE

Input: $P = \{x_i\}_{i=1}^n \sim Exp(\lambda)$, bounds $0 < \lambda_{\min} < \lambda_{\max}$, target error $\alpha \in (0,1)$, privacy budget $\epsilon > 0$.
Output: Parameter estimator $\tilde{\lambda}$.
procedure MLE-LEARNING($P, \lambda_{\min}, \lambda_{\max}, \alpha, \epsilon$)
  $\tilde{R} \leftarrow$ QUANTILE-ESTIMATION($P, \lambda_{\min}, \lambda_{\max}, \theta = 1/10, \epsilon/2$) ▷ Range estimation
  $\tilde{\lambda} \leftarrow$ PRIVATE-MLE($P, \tilde{R}\log(n), \epsilon/2$)
  return $\tilde{\lambda}$

Theorem 3.6. *Algorithm 3 is ϵ-differentially private, and w.p. $\geq 1-\beta$ it returns a $(1\pm\alpha)$-approximation of λ, given that:*

$$n \;\geq\; O\left(\max\left\{\frac{\ln(\frac{1}{\alpha})\ln(\frac{1}{\beta})\ln(n)}{\lambda\epsilon\alpha}, \frac{\ln(\frac{1}{\beta})}{\alpha^2}, \frac{\ln(\frac{\ln(\lambda_{\max}/\lambda_{\min})}{\beta})}{\epsilon}\right)\right\}$$

Proof. Privacy. The algorithm is a sequential composition of the quantile/range subroutine run with privacy budget $\epsilon/2$ (Theorem 3.1) and the private MLE subroutine run with privacy budget $\epsilon/2$ (Theorem 3.4). By basic composition, the overall privacy loss is at most ϵ.

[1] $\alpha = O(1)$ and $e^{-\lambda R} = O(1/poly(n))$

Utility. With the choice $\theta = 1/10$ in the range routine and by its utility guarantee (Theorem 3.2), with probability $\geq 1-\beta/2$ we obtain $\tilde{Q}_{1-\theta}$ within a constant factor of $Q_{1-\theta} = \frac{1}{\lambda}\ln(1/\theta)$, and hence (by Theorem 3.3) a clipping level $\Theta\Big(\frac{1}{\lambda}\cdot\ln n\Big)$. Consequently $e^{-\lambda\tilde{R}} = O(1/poly(n)) \leq \alpha/2$. Plugging this $\tilde{R}$ into the MLE analysis (Theorem 3.5 with budget $\epsilon/2$ and confidence $\beta/2$) yields the utility of Algorithm 3. Union-bounding the two stages gives success probability $\geq 1-\beta$. Collecting the dominant terms proves the stated big-O bound. □

3.2 Learning λ from the $(1-1/e)$-Quantile

We now present a simplified approach for privately estimating λ that bypasses range estimation and MLE entirely. The method relies on the fact that the $(1-1/e)$-quantile of $\mathrm{Exp}(\lambda)$ equals $Q_{1-1/e} = \frac{1}{\lambda}$. Thus, a multiplicative $(1\pm\alpha)$-approximation to $Q_{1-1/e}$ directly yields a multiplicative $(1\pm\alpha)$-approximation to $1/\lambda$, and therefore to λ itself. We use a binary search on powers of $(\frac{1}{1-\alpha/2})$ that lie inside the interval $[\lambda_{\min},\lambda_{\max}]$ for known bounds $0<\lambda_{\min}<\lambda_{\max}$, with the comparison at each step implemented via a private counting query. The output of the search is $\tilde{q}$, which we interpret as $\tilde{q}\approx\frac{1}{\lambda}$, so our final estimator is $\tilde{\lambda}=\frac{1}{\tilde{q}}$.

Algorithm 4 Private Exponential Learning via $1/e$-Quantile

Input: $P=\{x_i\}_{i=1}^n \sim \mathrm{Exp}(\lambda)$, bounds $0<\lambda_{\min}<\lambda_{\max}$, target error $\alpha\in(0,1)$, privacy budget $\epsilon>0$.
Output: Parameter estimator $\tilde{\lambda}$.
procedure QUANTILE-LEARNING($P,\lambda_{\min},\lambda_{\max},\alpha,\epsilon$)
  Denote $[p_1=\lambda_{\min}, p_2=\lambda_{\min}(\frac{1}{1-\alpha/2}), p_3=\lambda_{\min}(\frac{1}{1-\alpha/2})^2,\ldots,p_{\lceil\log_2(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\alpha})\rceil}=\lambda_{\max}]$
  Denote $L\leftarrow 1$, $U\leftarrow\left\lceil\log_2\Big(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\alpha}\Big)\right\rceil$
  for $t=1$ **to** $\left\lceil\log_2\Big(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\alpha}\Big)\right\rceil$ **do**
    $m\leftarrow\frac{U+L}{2}$
    $\tilde{q}\leftarrow\frac{1}{n}|\{x_i\in P: x_i<p_m\}| + \mathrm{Lap}\left(\frac{\left\lceil\log_2\Big(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\alpha}\Big)\right\rceil}{\epsilon n}\right)$
    if $\tilde{q}\geq 1-1/e+\alpha/2e$ **then**
      $U\leftarrow m$
    else if $\tilde{q}\leq 1-1/e-\alpha/2e$ **then**
      $L\leftarrow m$
    else
      return $\tilde{\lambda}=1/\tilde{q}$
  return $\perp$

Lemma 3.7 (Privacy of Algorithm 4)**.** Algorithm 4 is ϵ-differentially private.

The proof is deferred to Theorem A.8.

Lemma 3.8 (Utility of Algorithm 4)**.** Fix target accuracy $\alpha\in(0,1)$ and failure probability $\beta\in(0,1)$. Let $T=\left\lceil\log_2\Big(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\alpha}\Big)\right\rceil$. If

$$n \;\geq\; \max\left\{\frac{2eT}{\epsilon\alpha}\ln\left(\frac{2T}{\beta}\right),\frac{2}{\alpha^2}\ln\left(\frac{2T}{\beta}\right)\right\}$$

then with probability at least $1-\beta$ Algorithm 4 outputs $\tilde{\lambda}$ satisfying $\tilde{\lambda}\in[(1-\alpha)\lambda,\ (1+\alpha)\lambda]$.

Proof sketch. The algorithm performs a binary search over $T=\lceil\log_2(\log(\lambda_{\max}/\lambda_{\min})/\alpha)\rceil$ checkpoints, where each step compares the (noisy) empirical CDF at a candidate p to $1-1/e$.

Sampling error. By DKW, $\Pr[\sup_x|F_n(x)-F(x)|>\alpha/2]\leq 2e^{-2n(\alpha/2)^2}$. A union bound over the T comparisons shows that taking $n\geq\frac{2}{\alpha^2}\ln\Big(\frac{2T}{\beta}\Big)$ ensures all empirical CDF values deviate from the truth by at most $\alpha/2$ with probability $\geq 1-\beta/2$.

Noise error. Each comparison releases one count with Laplace noise of scale $b=T/(\epsilon n)$. Requiring that every noisy perturbation be at most $\alpha/(2e)$ (so that the CDF-side error budget totals $\alpha/2$ from privacy) holds with probability $\geq 1-\beta/2$ if $n\geq\frac{2eT}{\epsilon\alpha}\ln\Big(\frac{2T}{\beta}\Big)$.

From CDF to parameter accuracy. For $X\sim\mathrm{Exp}(\lambda)$, the $(1-1/e)$-quantile equals $1/\lambda$. An additive CDF error of at most $\alpha/(2e)$ around the level $1-1/e$ corresponds to at least $\alpha/(2e)$ probability mass in the exponential distribution. This amount of probability mass separates the true quantile $1/\lambda$ from the neighboring points $1/((1\pm\alpha)\lambda)$. Therefore, controlling the CDF error at this level guarantees that the estimated quantile lies in $\left[\frac{1}{(1+\alpha)\lambda},\frac{1}{(1-\alpha)\lambda}\right]$, and thus its reciprocal $\tilde{\lambda}$ falls in $[(1-\alpha)\lambda,(1+\alpha)\lambda]$.

Combining all requirements via the union bound yields the stated sample-size condition with success probability $1-\beta$. Full details are deferred to Theorem A.9. □

The following theorem summarizes the properties of Algorithm 4.

Theorem 3.9. *Algorithm 4 is ϵ-differentially private and, for any $\alpha\in(0,1)$ and $\beta\in(0,1)$, achieves an $(1\pm\alpha)$-approximation to λ with probability at least $1-\beta$, provided*

$$n \;\geq\; O\left(\max\left\{\frac{\ln(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\alpha})}{\epsilon\alpha}\ln\Big(\frac{\log(\lambda_{\max}/\lambda_{\min})/\alpha}{\beta}\Big),\ \frac{1}{\alpha^2}\ln\Big(\frac{\log(\lambda_{\max}/\lambda_{\min})/\alpha}{\beta}\Big)\right\}\right)$$

Proof. Privacy follows from Theorem 3.7. Utility follows from Theorem 3.8. □

3.3 Best of Both Algorithms

We now combine the two estimators described above—the MLE-based approach (Section 3.1, Algorithm 3) and the $(1-1/e)$-quantile approach (Section 3.2, Algorithm 4)—into a single algorithm that chooses the better option based on the (private) magnitude of λ. As shown above, the resp. sample complexities are

$$\text{MLE:} \quad \tilde{O}\left(\max\left\{\tfrac{1}{\lambda\epsilon\alpha},\ \tfrac{1}{\epsilon},\ \tfrac{1}{\alpha^2}\right\}\right)$$

$$(1-1/e)\text{-quantile:} \quad \tilde{O}\left(\max\left\{\tfrac{1}{\epsilon\alpha},\ \tfrac{1}{\alpha^2}\right\}\right)$$

so the only asymptotic difference is between $\frac{1}{\lambda\epsilon\alpha}$ and $\frac{1}{\epsilon\alpha}$. Hence, when $\lambda > 1$ the MLE approach is better, and when $\lambda \le 1$ the $(1-1/e)$-quantile approach is better.

Algorithm 5 Adaptive Private Exponential Learning (Best-of-Both)

Input: $P = \{x_i\}_{i=1}^n$, bounds $0 < \lambda_{\min} < \lambda_{\max}$, target error $\alpha \in (0,1)$, privacy budget $\epsilon > 0$.
Output: $\tilde{\lambda}$.
procedure BEST-OF-BOTH($P, \lambda_{\min}, \lambda_{\max}, \alpha, \epsilon$)
  $\tilde{\lambda}_0 \leftarrow$ QUANTILE-LEARNING($P,\ \lambda_{\min},\ \lambda_{\max}, \alpha = 1/2,\ \epsilon/3$) (Alg. 4)
  if $\tilde{\lambda}_0 \ge 2$ **then** ▷ confidently "large"
    $\tilde{\lambda} \leftarrow$ MLE-LEARNING($P,\ \lambda_{\min},\ \lambda_{\max},\ \alpha,\ 2\epsilon/3$) (Alg. 3)
  else
    $\tilde{\lambda} \leftarrow$ QUANTILE-LEARNING($P,\ \lambda_{\min},\ \lambda_{\max},\ \alpha,\ 2\epsilon/3$) (Alg. 4)
  return $\tilde{\lambda}$

Algorithm 4 can produce a constant-factor estimate $\tilde{\lambda}_0$ of λ using $O(1)$ accuracy. With such a coarse estimate, the events $\tilde{\lambda}_0 \ge 2$ and $\tilde{\lambda}_0 \le \frac{1}{2}$ imply, respectively, that $\lambda > 1$ and $\lambda < 1$. The middle region $\tilde{\lambda}_0 \in (\frac{1}{2}, 2)$ corresponds to $\lambda \in (\frac{1}{2}, 2)$, where both routes have linear terms within a factor at most 2, and either choice is acceptable up to constants; we default to the quantile route.

Lemma 3.10 (Privacy of Alg. 5)**.** Algorithm 5 is ϵ-differentially private.

The proof is deferred to Theorem A.10.

Theorem 3.11. *Fix $\alpha \in (0,1)$, $\beta \in (0,1)$ and $\epsilon > 0$. Algorithm 5 is ϵ-differentially private and, with probability at least $1-\beta$, outputs $\tilde{\lambda}$ satisfying $\tilde{\lambda} \in [(1-\alpha)\lambda, (1+\alpha)\lambda]$, provided that*

$$n \ge \tilde{O}\left(\max\left\{\min\left\{\frac{1}{\epsilon\alpha}, \frac{1}{\lambda\epsilon\alpha}\right\}, \frac{1}{\alpha^2}, \frac{1}{\epsilon}\right\}\right)$$

Equivalently, up to polylogarithmic factors, the adaptive procedure matches the best of the two options:

$$\text{if } \lambda \ge 1 \quad \Rightarrow \quad \tilde{O}\left(\max\left\{\tfrac{1}{\lambda\epsilon\alpha},\ \tfrac{1}{\epsilon},\ \tfrac{1}{\alpha^2}\right\}\right),$$

$$\text{if } \lambda \le 1 \quad \Rightarrow \quad \tilde{O}\left(\max\left\{\tfrac{1}{\epsilon\alpha},\ \tfrac{1}{\alpha^2}\right\}\right)$$

Proof. Privacy follows from Theorem 3.10. For *utility*, let $\mathcal{E}_{\text{est}}$ be the event that the coarse step attains a constant-factor estimate with probability at least $1 - \beta/3$ (Theorem 3.8 with target error $= \frac{1}{2}$ and privacy budget $\epsilon/3$). Under $\mathcal{E}_{\text{est}}$, the decision matches the oracle choice whenever $\lambda \notin (\frac{1}{2}, 2)$, and in the middle region the two routes are within a constant factor anyway. Conditioned on the chosen branch, the accuracy $\tilde{\lambda} \in [(1-\alpha)\lambda, (1+\alpha)\lambda]$ holds with probability at least $1 - 2\beta/3$ by Theorem 3.5 for the MLE route (using budget $2\epsilon/3$) or by Theorem 3.8 for the quantile route (again with budget $2\epsilon/3$). A union bound over the two stages yields overall success probability at least $1-\beta$. The stated sample-size bound aggregates the dominating terms of the chosen branch. □

4 Applications to Pareto Distributions

The exponential distribution is not only fundamental in its own right but also serves as a key building block for other families. In particular, the *Pareto distribution*—one of the most widely studied heavy-tailed laws in economics, finance, and network modeling—has probability density

$$f(x) = \alpha_p \cdot \frac{x_m^{\alpha_p}}{x^{\alpha_p+1}}, \qquad x \ge x_m > 0,$$

where x_m is the scale parameter (minimum value) and α_p is the shape parameter. Through the logarithmic transform $y = \ln(x/x_m)$, a Pareto(x_m, α_p) random variable reduces exactly to Exp(α_p). This reduction allows us to transfer our private exponential learning algorithms directly to the Pareto setting.

A subtlety arises, however, in the case when x_m is unknown: direct release of the sample minimum is infeasible under differential privacy due to its high sensitivity. Instead, one must combine quantile-based estimation with the logarithmic reduction in order to privately and accurately recover both (α_p, x_m).

We develop and analyze these reductions formally, showing how our best-of-both exponential learner extends to Pareto distributions while preserving privacy and accuracy. The full algorithms, proofs, and utility guarantees are deferred to Section B.

5 Lower Bound for Exponential Distribution Learning under Pure Differential Privacy

To complement our algorithmic results, we establish lower bounds on the sample complexity of privately learning exponential distributions under pure DP. Our construction follows a standard packing argument: we identify a family of rate parameters Λ that are well-separated in TV-distance, and then invoke DP generalization bounds to show that distinguishing them requires a large number of samples. Our lower bound is based on the group privacy approach of Karwa and Vadhan (2017), which provides a powerful technique for deriving lower bounds in private distribution learning. This demonstrates that our upper bounds from Section 3 are essentially tight, up to logarithmic factors.

Denote Λ as the set

$$\Lambda = \left\{\lambda_{\min}, \lambda_{\min}(1+8\alpha), \lambda_{\min}(1+8\alpha)^2, \ldots, \lambda_{\max}\right\}$$

and consider the following family of $t = |\Lambda|$ distributions:

$$\{P_i = \mathrm{Exp}(\lambda_i) : \ \lambda_i \in \Lambda\}$$

so the rate parameters range geometrically from $\lambda_{\min}$ up to $\lambda_{\max}$ with factor $(1+8\alpha)$.

We first show that $\forall i \neq j : \ TV(P_i, P_j) > \alpha$. Fix $i > j$ so that $\lambda_i > \lambda_j$. Then, by Theorem 2.8:

$$TV(\mathrm{Exp}(\lambda_i), \mathrm{Exp}(\lambda_j)) = e^{-\lambda_j \cdot a} - e^{-\lambda_i \cdot a}$$

where

$$a = \frac{\ln(\lambda_i) - \ln(\lambda_j)}{\lambda_i - \lambda_j} = \frac{\ln(\frac{\lambda_i}{\lambda_j})}{\lambda_i - \lambda_j}$$

Then we get:

$$\begin{aligned} TV(\mathrm{Exp}(\lambda_i), \mathrm{Exp}(\lambda_j)) &= e^{-\lambda_i \cdot \frac{\ln((\lambda_i/\lambda_j)}{\lambda_i - \lambda_j}} - e^{-\lambda_j \cdot \frac{\ln(\lambda_i/\lambda_j)}{\lambda_i - \lambda_j}} \\ &= e^{-\frac{\ln(\lambda_i/\lambda_j)}{(\lambda_i/\lambda_j)-1}} - e^{-(\lambda_i/\lambda_j)\frac{\ln(\lambda_i/\lambda_j)}{(\lambda_i/\lambda_j)-1}} \overset{r=\lambda_i/\lambda_j}{=} r^{-\frac{1}{r-1}} - r^{-\frac{r}{r-1}} \\ &= r^{-\frac{1}{r-1}} \cdot (1 - r^{-1}) \end{aligned}$$

Now, the following lemma proves that $TV(P_i, P_j) > \alpha$.

Lemma 5.1. Let $T(r)$ denote the total variation distance between two exponential distributions whose rates differ by a ratio $r \geq 1$:

$$\begin{aligned} T(r) &:= \mathrm{TV}\big(\mathrm{Exp}(r\lambda), \mathrm{Exp}(\lambda)\big) = r^{-\frac{1}{r-1}} - r^{-\frac{r}{r-1}} \\ &= r^{-\frac{1}{r-1}}\big(1 - \tfrac{1}{r}\big) \end{aligned}$$

Fix any $\alpha \in (0, \frac{1}{2})$. Then there exists an absolute constant c such that, with $r = 1 + c\alpha$, one has $T(r) \geq \alpha$. In particular, the choice $c = 8$ suffices.

The proof is deferred to Theorem A.12. Now we can state the following lemma.

Lemma 5.2 (Formal lower bound)**.** Fix parameters $0 < \alpha < \frac{1}{2}$, $0 < \beta < \frac{1}{2}$, and an interval $[\lambda_{\min}, \lambda_{\max}]$ with $0 < \lambda_{\min} < \lambda_{\max}$. Let $\mathcal{M}$ be any ϵ-differentially private algorithm that, given n i.i.d. samples from $\mathrm{Exp}(\lambda)$, outputs $\tilde{\lambda}$ such that $\Pr[\,\tilde{\lambda} \in [(1-\alpha)\lambda, (1+\alpha)\lambda]\,] \geq 1 - \beta$ for every $\lambda \in [\lambda_{\min}, \lambda_{\max}]$. Then there exists a universal constant $c > 0$ such that the sample size must satisfy

$$n \ \geq \ \frac{c}{\epsilon\,\alpha} \ \ln\left(\frac{\frac{\log(\lambda_{\max}/\lambda_{\min})}{\alpha}}{\beta}\right)$$

Corollary 5.3. Any ϵ-DP learner achieving TV-error at most α (via Corollary 2.9) also requires $n = \Omega\big(\frac{1}{\epsilon\alpha} \ln(\frac{\ln(\lambda_{\max}/\lambda_{\min})/\alpha}{\beta})\big)$.

Proof of Theorem 5.2. Fix i and let $\tilde{\lambda}$ denote the output of any private exponential distribution learner. Then, by construction, for all $j \neq i$,

$$\Pr_{P \sim P_i^n}\left[\tilde{\lambda} \in (1 \pm \alpha)\tfrac{1}{(1+8\alpha)^j}\right] \ \leq \ \beta$$

In contrast, we can use the following lemma.

Lemma 5.4 (Lemma 6.1 of Karwa and Vadhan (2017))**.** For every pair of distributions $\mathbb{D}_{\theta_0}, \mathbb{D}_{\theta_1}$, every (ϵ, δ)-differentially private algorithm $M(x_1, \ldots, x_n)$, if $M_{\theta_0}, M_{\theta_1}$ are two induced marginal distributions on the output of M evaluated on input dataset $X_1, \ldots, X_n$ sampled i.i.d. from $\mathbb{D}_{\theta_0}$ and $\mathbb{D}_{\theta_1}$ respectively, $\epsilon' = 6\epsilon n \|\mathbb{D}_{\theta_0} - \mathbb{D}_{\theta_1}\|_{tv}$ and $\delta' = 4e^{\epsilon'} n\delta \|\mathbb{D}_{\theta_0} - \mathbb{D}_{\theta_1}\|_{tv}$, then, for every event E,

$$M_{\theta_0}(E) \leq e^{\epsilon'} M_{\theta_1}(E) + \delta'$$

The lemma above implies

$$\begin{aligned} &\Pr_{P \sim P_i^n}\left[\tilde{\lambda} \notin (1 \pm \alpha)\tfrac{1}{(1+8\alpha)^i}\right] \\ &\geq \sum_{\substack{j=1 \\ j \neq i}}^{t} \Pr_{P \sim P_i^n}\left[\tilde{\lambda} \in (1 \pm \alpha)\tfrac{1}{(1+8\alpha)^j}\right] \\ &\geq \sum_{\substack{j=1 \\ j \neq i}}^{t} e^{6n\epsilon\,\mathrm{TV}(P_i, P_j)} \Pr_{P \sim P_j^n}\left[\tilde{\lambda} \in (1 \pm \alpha)\tfrac{1}{(1+8\alpha)^j}\right] \\ &\geq \ t \cdot e^{6n\epsilon\alpha}(1-\beta) \ \overset{\beta \leq 1/2}{\geq} \ \tfrac{1}{2} t \cdot e^{6n\epsilon\alpha} \end{aligned}$$

Noting that $\lambda_{\min}(1+8\alpha)^t = \lambda_{\max}$ we get that

$$t = \frac{\log(\lambda_{\max}/\lambda_{\min})}{\log(1+8\alpha)} \geq \frac{1}{8\alpha}\log(\lambda_{\max}/\lambda_{\min})$$

we conclude that $\frac{1}{16\alpha}\ln(\lambda_{\max}/\lambda_{\min})\cdot e^{6n\epsilon\alpha} \le \beta$, which rearranges to the following lower bound on the required sample size:

$$n \ge \frac{1}{6\epsilon\alpha}\ln\left(\frac{\frac{\ln(\lambda_{\max}/\lambda_{\min})}{16\alpha}}{\beta}\right) \qquad \square$$

6 Finding Initial Bounds Using Approximate Differential Privacy

The algorithms of Section 3 required prior bounds $(\lambda_{\min}, \lambda_{\max})$ on the range of plausible λ in order to calibrate privacy noise. As shown by our lower bound in Section 5, such knowledge of $(\lambda_{\min}, \lambda_{\max})$ is in fact *necessary* for pure differential privacy, since otherwise no algorithm can achieve nontrivial accuracy. In this section we show that under approximate (ϵ, δ)-DP we can eliminate this requirement: using a median-based private routine, we can obtain valid initial bounds directly from the data, and then feed them into our algorithm. Thus the extra logarithmic dependence on $\lambda_{\max}/\lambda_{\min}$ is an artifact of pure DP, not of the problem itself.

To eliminate the need for externally provided bounds $(\lambda_{\min}, \lambda_{\max})$ under (ϵ, δ)-DP, we design a private subroutine that learns such bounds directly from the data. The key idea is to approximate the median of the exponential distribution using noisy counts over dyadic intervals; since the median is tightly concentrated around $\ln(2)/\lambda$, this gives a natural anchor point. By expanding outward from this approximate median, we obtain lower and upper bounds $\tilde{\lambda}_{\min}$ and $\tilde{\lambda}_{\max}$ that safely contain the true median and hence allow us to calibrate the later learning algorithms. This approach is based on the differentially private histogram technique of Vadhan (2017), which provides an efficient way to identify approximate median under privacy constraints. The procedure is summarized in Algorithm 6 below.

Lemma 6.1. Algorithm 6 is (ϵ, δ)-differentially private. Moreover, if

$$n \ge \max\left\{\frac{800}{\epsilon}\ln(\frac{2}{\delta\beta}), 5000\ln(\frac{2}{\beta})\right\}$$

then

$$\Pr[\tilde{\lambda}_{\min} < \lambda < \tilde{\lambda}_{\max}] \ge 1-\beta$$

The proof is deferred to Theorem A.11.

We now state the main guarantee of Algorithm 6.

Theorem 6.2. *There exists an (ϵ, δ)-differentially private algorithm that, for any $\alpha \in (0,1)$ and $\beta \in (0,1)$, achieves a $(1 \pm \alpha)$-approximation to λ with probability at least $1-\beta$, provided*

$$n \ge O\left(\max\left\{\min\left\{\frac{\ln(1/\alpha)}{\epsilon\alpha}\ln\left(\frac{\ln(1/\alpha)}{\beta}\right), \frac{\ln(1/\alpha)\ln(1/\beta)\ln(n)}{\lambda\epsilon\alpha}\right\}, \right.\right.$$
$$\left.\left.\frac{1}{\alpha^2}\ln\left(\frac{\ln(1/\alpha)}{\beta}\right), \frac{\ln(1/\beta)}{\alpha^2}, \frac{\ln(1/\delta\beta)}{\epsilon}\right\}\right)$$

Algorithm 6 Finding initial private approximated bounds $0 < \tilde{\lambda}_{\min} < \tilde{\lambda}_{\max}$

Input: $P = \{x_i\}_{i=1}^n \sim Exp(\lambda)$, privacy budget ϵ, δ.
Output: bounds $0 < \tilde{\lambda}_{\min} < \tilde{\lambda}_{\max}$.
For all $k \in \mathbb{Z}$, set $c_k(P) = \frac{1}{n}|\{x_i \mid x_i \in (2^k, 2^{k+1})\}|$.
For all $c_k(P)$, set $\hat{c}_k(P) = 0$ if $c_k(P) = 0$. Else set $\hat{c}_k(P) = c_k(P) + Z_k$, where each $Z_k \leftarrow \mathrm{Lap}(2/\epsilon n)$ independently.
Set $S = \{k \mid \hat{c}_k(P) \ge \frac{2}{\epsilon n}\log(2/\delta) + \frac{1}{n}\}$.
if $|S| \ge 1$ **then**
  Let $k^* \leftarrow \arg\max_{k\in S} \hat{c}_k(P)$
  return $\tilde{\lambda}_{\min} = \ln(2)(2^{k^*+1})^{-1}$, $\tilde{\lambda}_{\max} = \ln(2)(2^{k^*-1})^{-1}$
else
  return $\perp$

7 Conclusion

We introduce the first differentially private algorithms for learning Exponential distributions and utilize them as building blocks for learning Pareto distributions. Our results reveal a novel phase transition in the role of the rate parameter λ: depending on its magnitude, either the MLE-based or the quantile-based approach is superior. We resolved this tension with an adaptive best-of-both algorithm that adaptively selects the better approach. Additionally, we proved lower bounds that match our upper bounds up to logarithmic factors, establishing near-tight sample complexity. We further showed that moving from pure to approximate differential privacy removes the dependency on externally provided bounds, underscoring a structural difference between the two regimes.

Looking ahead, several directions remain open. Can our techniques be extended to multi-parameter exponential families or to mixtures? What are the precise tradeoffs for other heavy-tailed distributions? And more broadly, can adaptive best-of-both strategies be systematically developed for other problems in private statistics? We hope this work motivates further exploration of learning additional heavy-tailed distributions under differential privacy.

References

Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. Differentially private testing of identity and close-

ness of discrete distributions, 2017. URL https://arxiv.org/abs/1707.05128.

Søren Asmussen. *Applied Probability and Queues*. IEEE Computer Society Press, United States, 1987. ISBN 0471911739.

Dimitri Bertsekas and Robert Gallager. *Data networks (2nd ed.)*. Prentice-Hall, Inc., USA, 1992. ISBN 0132009161.

Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In Martin Hirt and Adam Smith, editors, *Theory of Cryptography*, pages 635–658, Berlin, Heidelberg, 2016. Springer Berlin Heidelberg. ISBN 978-3-662-53641-4.

Mark Bun, Jonathan Ullman, and Salil Vadhan. Fingerprinting codes and the price of approximate differential privacy. In *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*, STOC '14, page 1–10, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450327107. doi: 10.1145/2591796.2591877. URL https://doi.org/10.1145/2591796.2591877.

Mark Bun, Kobbi Nissim, Uri Stemmer, and Salil Vadhan. Differentially private release and learning of threshold functions. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 634–649, 2015. doi: 10.1109/FOCS.2015.45.

M.E. Crovella and A. Bestavros. Self-similarity in world wide web traffic: evidence and possible causes. *IEEE/ACM Transactions on Networking*, 5(6):835–846, 1997. doi: 10.1109/90.650143.

Mark S. Daskin. *Fundamentals of Queueing Theory*, pages 119–123. Springer International Publishing, Cham, 2021. ISBN 978-3-031-02493-1. doi: 10.1007/978-3-031-02493-1_17. URL https://doi.org/10.1007/978-3-031-02493-1_17.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In Shai Halevi and Tal Rabin, editors, *Theory of Cryptography*, pages 265–284, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg. ISBN 978-3-540-32732-5.

Anja Feldmann, Anna C. Gilbert, Polly Huang, and Walter Willinger. Dynamics of ip traffic: a study of the role of variability and the impact of control. *SIGCOMM Comput. Commun. Rev.*, 29(4):301–313, August 1999. ISSN 0146-4833. doi: 10.1145/316194.316235. URL https://doi.org/10.1145/316194.316235.

Xavier Gabaix. Power laws in economics: An introduction. *The Journal of Economic Perspectives*, 30(1):185–205, 2016. ISSN 08953309. URL http://www.jstor.org/stable/43710016.

Gordon Johnston. Statistical models and methods for lifetime data. *Technometrics*, 45:264–265, 01 2012. doi: 10.1198/tech.2003.s767.

Gautam Kamath, Jerry Li, Vikrant Singhal, and Jonathan Ullman. Privately learning high-dimensional distributions, 2019. URL https://arxiv.org/abs/1805.00216.

Gautam Kamath, Vikrant Singhal, and Jonathan Ullman. Private mean estimation of heavy-tailed distributions, 2021. URL https://arxiv.org/abs/2002.09464.

Vishesh Karwa and Salil P. Vadhan. Finite sample differentially private confidence intervals. In *Information Technology Convergence and Services*, 2017. URL https://api.semanticscholar.org/CorpusID:4670356.

Daniel Kifer, Adam Smith, and Abhradeep Thakurta. Private convex empirical risk minimization and high-dimensional regression. In Shie Mannor, Nathan Srebro, and Robert C. Williamson, editors, *Proceedings of the 25th Annual Conference on Learning Theory*, volume 23 of *Proceedings of Machine Learning Research*, pages 25.1–25.40, Edinburgh, Scotland, 25–27 Jun 2012. PMLR. URL https://proceedings.mlr.press/v23/kifer12.html.

Leonard Kleinrock. *Theory, Volume 1, Queueing Systems*. Wiley-Interscience, USA, 1975. ISBN 0471491101.

Benoit Mandelbrot. The pareto-lévy law and the distribution of income. *International Economic Review*, 1(2):79–106, 1960. ISSN 00206598, 14682354. URL http://www.jstor.org/stable/2525289.

S.M. Ross. *Stochastic processes*. Wiley series in probability and statistics: Probability and statistics. Wiley, 1996. ISBN 9780471120629. URL https://books.google.de/books?id=ImUPAQAAMAAJ.

Adam Smith. Privacy-preserving statistical estimation with optimal convergence rates. In *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing*, STOC '11, page 813–822, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450306911. doi: 10.1145/1993636.1993743. URL https://doi.org/10.1145/1993636.1993743.

Donald Smith. Statistical theory of reliability and life testing–probability models. *Technometrics*, 19, 04 2012. doi: 10.1080/00401706.1977.10489538.

Salil Vadhan. *The Complexity of Differential Privacy*, pages 347–450. Springer International Publishing, Cham, 2017. ISBN 978-3-319-57048-8. doi: 10.1007/978-3-319-57048-8_7. URL https://doi.org/10.1007/978-3-319-57048-8_7.

A Missing Proofs

Fact A.1 (Theorem 2.8 restated)**.** Let $\mathrm{Exp}(\lambda_1), \mathrm{Exp}(\lambda_2)$ be two exponential distributions and assume $\lambda_1 \geq \lambda_2$. Denote $a = \frac{\ln(\lambda_1)-\ln(\lambda_2)}{\lambda_1-\lambda_2}$, then:

$$TV(\mathrm{Exp}(\lambda_1), \mathrm{Exp}(\lambda_2)) = e^{-\lambda_2 \cdot a} - e^{-\lambda_1 \cdot a}$$

Proof. By definition of total variation and the PDF of exponential distribution:

$$TV(\mathrm{Exp}(\lambda_1), \mathrm{Exp}(\lambda_2)) = \frac{1}{2}\int_0^\infty |\lambda_1 e^{-\lambda_1 x} - \lambda_2 e^{-\lambda_2 x}|\, dx$$

Recall that $\lambda_1 \geq \lambda_2$ without loss of generality, then let's compute the unique point where the two PDFs meet:

$$\lambda_1 e^{-\lambda_1 a} = \lambda_2 e^{-\lambda_2 a} \iff \ln(\lambda_1) - \lambda_1 a = \ln(\lambda_2) - \lambda_2 a \iff a = \frac{\ln(\lambda_1) - \ln(\lambda_2)}{\lambda_1 - \lambda_2}$$

Now we compute:

$$\begin{aligned}\int_0^\infty |\lambda_1 e^{-\lambda_1 x} - \lambda_2 e^{-\lambda_2 x}|\, dx &= \int_0^a (\lambda_1 e^{-\lambda_1 x} - \lambda_2 e^{-\lambda_2 x})\, dx \; - \int_a^\infty (\lambda_1 e^{-\lambda_1 x} - \lambda_2 e^{-\lambda_2 x})\, dx \\ &= (e^{-\lambda_2 x} - e^{-\lambda_1 x})\big|_0^a - (e^{-\lambda_2 x} - e^{-\lambda_1 x})\big|_a^\infty = 2(e^{-\lambda_2 a} - e^{-\lambda_1 a})\end{aligned}$$

Hence we get:

$$TV(\mathrm{Exp}(\lambda_1), \mathrm{Exp}(\lambda_2)) = e^{-\lambda_2 \cdot a} - e^{-\lambda_1 \cdot a}$$

□

Corollary A.2 (Theorem 2.9 restated)**.** If the rate parameters satisfy $\lambda_1 \in (1 \pm \alpha)\lambda_2$, then

$$TV(\mathrm{Exp}(\lambda_1), \mathrm{Exp}(\lambda_2)) \leq \alpha$$

Proof. Assume without loss of generality that $\lambda_1 \geq \lambda_2$. We first consider the case where $\lambda_1 = (1+\delta)\lambda_2$ with $|\delta| \leq \alpha$. From the fact above, the total variation distance is:

$$TV(\mathrm{Exp}(\lambda_1), \mathrm{Exp}(\lambda_2)) = e^{-\lambda_2 a} - e^{-\lambda_1 a}$$

where

$$a = \frac{\ln(\lambda_1) - \ln(\lambda_2)}{\lambda_1 - \lambda_2} = \frac{\ln(1+\delta)}{\delta\lambda_2}$$

Substituting into the formula, we get:

$$\begin{aligned}TV(\mathrm{Exp}(\lambda_1), \mathrm{Exp}(\lambda_2)) &= e^{-\lambda_2 \cdot \frac{\ln(1+\delta)}{\delta\lambda_2}} - e^{-(1+\delta)\lambda_2 \cdot \frac{\ln(1+\delta)}{\delta\lambda_2}} \\ &= e^{-\frac{\ln(1+\delta)}{\delta}} - e^{-(1+\delta)\frac{\ln(1+\delta)}{\delta}} \\ &= (1+\delta)^{-1/\delta} - (1+\delta)^{-(1+\delta)/\delta} \\ &= (1+\delta)^{-1/\delta} \cdot (1 - \frac{1}{(1+\delta)})\end{aligned}$$

Since $\delta \leq \alpha$ and the expression is monotonically increasing in $\delta > 0$, we can upper bound:

$$TV(\mathrm{Exp}(\lambda_1), \mathrm{Exp}(\lambda_2)) \leq (1+\alpha)^{-1/\alpha} \cdot (1 - \frac{1}{(1+\alpha)})$$

Finally

$$(1+\alpha)^{1/\alpha} \cdot (1 - \frac{1}{(1+\alpha)}) = \alpha \cdot (1+\alpha)^{-(1+\alpha)/\alpha}$$

Now compute:

$$(1+\alpha)^{-(1+\alpha)/\alpha} = e^{\ln((1+\alpha)^{-(1+\alpha)\alpha})} = e^{-\frac{1+\alpha}{\alpha}\ln(1+\alpha)}$$
$$\overset{(*)}{\leq} e^{-\frac{1+\alpha}{\alpha}\cdot\frac{\alpha}{1+\alpha}} = e^{-1} < 1$$

where inequality $(*)$ follows from $\ln(1+\alpha) \geq \frac{\alpha}{1+\alpha}$.

Now consider $\lambda_2 = (1-\delta)\lambda_1$ with $|\delta| \leq \alpha$. then:

$$a = \frac{-\ln(1-\delta)}{\delta\lambda_1}$$

Following the same way, substituting into the formula, we get:

$$\begin{aligned} TV(\mathrm{Exp}(\lambda_1), \mathrm{Exp}(\lambda_2)) &= e^{(1-\delta)\lambda_1 \cdot \frac{\ln(1-\delta)}{\delta\lambda_1}} - e^{\lambda_1 \cdot \frac{\ln(1-\delta)}{\delta\lambda_1}} \\ &= e^{(1-\delta)\frac{\ln(1-\delta)}{\delta}} - e^{\frac{\ln(1-\delta)}{\delta}} \\ &= (1-\delta)^{(1-\delta)/\delta} - (1-\delta)^{1/\delta} \\ &= (1-\delta)^{1/\delta} \cdot (\frac{1}{1-\delta} - 1) \end{aligned}$$

Since $\delta \leq \alpha$ and the expression is monotonically decreasing in $\delta > 0$, we can upper bound:

$$\begin{aligned} TV(\mathrm{Exp}(\lambda_1), \mathrm{Exp}(\lambda_2)) &\leq \lim_{\delta\to 0}(1-\delta)^{1/\delta} \cdot (\frac{1}{1-\delta} - 1) \\ &= \lim_{\delta\to 0}(1-\delta)^{(1-\delta)/\delta} \cdot \delta = e^{-1} \cdot \delta \leq \alpha \end{aligned}$$

Overall we conclude:

$$TV(\mathrm{Exp}(\lambda_1), \mathrm{Exp}(\lambda_2)) \leq \alpha.$$

□

Lemma A.3 (Theorem 3.1 restated)**.** Algorithm 1 is ϵ-differentially private.

Proof. Let neighboring datasets P, P' differ in one record. For each $i \geq 0$ define the 1/n–sensitive counting query

$$q_i(P) \;=\; \frac{1}{n}\left|\{x \in P:\; x < 2^i\lambda_{\min}\}\right| \qquad (\Delta = 1/n)$$

The algorithm draws a noisy threshold $\tilde{T} = (1-\theta) + \mathrm{Lap}(2\Delta/\epsilon)$ and, scanning $i = 0, 1, 2, \ldots$, draws independent noisy query values $\tilde{q}_i = q_i(P) + \mathrm{Lap}(4\Delta/\epsilon)$ and halts at the first index with $\tilde{q}_i \geq \tilde{T}$, returning that index (equivalently, $2^i\lambda_{\min}$), or $\perp$ if none occurs.

This is exactly the *AboveThreshold* (a.k.a. the *Sparse Vector Technique*) with sensitivity $\Delta = 1/n$, a single positive report, and noise scales $2\Delta/\epsilon$ for the threshold and $4\Delta/\epsilon$ for the queries. By the standard SVT privacy guarantee, the mechanism that outputs only the first index i whose noisy query exceeds the noisy threshold is ϵ-differentially private, independent of the number of queries evaluated (see, e.g., Dwork et al., 2006, Theorem 3.23). The final reported value $2^i\lambda_{\min}$ is a deterministic post-processing of this index and therefore preserves ϵ-DP. □

Lemma A.4 (Theorem 3.2 restated)**.** Algorithm 1 returns a 6-approximation of the $(1-\theta)$-quantile of P w.p. $\geq 1-\beta$, given that $1/10 \leq \theta \leq 9/10$ and:

$$n \geq \max\left\{\frac{20}{\epsilon}\ln\left(\frac{4\log\left(\lambda_{\max}/\lambda_{\min}\right)}{\beta}\right), 200\ln\left(4/\beta\right)\right\}$$

Proof. The algorithm performs a search over dyadic multiples of the input parameter $\lambda_{\min}$, and outputs the smallest value $2^i \cdot \lambda_{\min}$ such that the noisy empirical cumulative distribution function (CDF) at that point exceeds a noisy threshold $\tilde{T} = (1-\theta) + \mathrm{Lap}(\frac{2}{\epsilon n})$.

Write $F(x) = 1 - e^{-\lambda x}$ for the true CDF and $F_n(x)$ for the empirical CDF. Two error sources affect each comparison with the threshold $1-\theta$:

1. **Sampling error.** By the Dvoretzky–Kiefer–Wolfowitz (DKW) inequality,

$$\Pr\Big[\sup_x \big|F_n(x) - F(x)\big| > \tfrac{1}{20}\Big] \;\le\; 2e^{-2n(1/20)^2}.$$

Requiring this to be $\le \beta/2$ yields

$$n \;\ge\; \frac{1}{2(1/20)^2}\ln\Big(\tfrac{4}{\beta}\Big) \;=\; 200\ln\Big(\tfrac{4}{\beta}\Big).$$

This implies that the algorithm's comparison of the empirical CDF to the threshold $(1-\theta)$ is accurate up to $1/20$ due to sampling error.

2. **Noise error.** Each comparison uses (i) a Laplace perturbation of the empirical CDF and (ii) a Laplace perturbation of the threshold. Let $b = 4/(\epsilon n)$ be the Laplace scale. For any $t > 0$, $\Pr\big[|\mathrm{Lap}(b)| > t\big] = e^{-t/b}$. Note that the threshold noise uses scale $2/(\epsilon n)$, which is smaller; thus, if we ensure that the Laplace noise with scale $4/(\epsilon n)$ is bounded, the threshold noise is automatically bounded as well. We require every noise draw among at most $2\log(\lambda_{\max}/\lambda_{\min})$ independent draws (one for the CDF at each of $\log(\lambda_{\max}/\lambda_{\min})$ checkpoints and one for the threshold; the factor 2 is a safe constant) to be bounded by $1/20$. By a union bound,

$$2\log(\lambda_{\max}/\lambda_{\min})\cdot e^{-(1/20)/b} \;\le\; \beta/2 \quad\Longleftrightarrow\quad n \;\ge\; \frac{5}{\epsilon}\ln\Big(\tfrac{4\log(\lambda_{\max}/\lambda_{\min})}{\beta}\Big).$$

Under these two events, we have with probability at least $1-\beta$, both the noisy empirical CDF and the threshold are within $1/20$ of their true values. Consequently, since $q_{1-\theta} = \frac{1}{\lambda}\ln(\frac{1}{\theta})$ for the exponential distribution, the decision to return a given $2^i\cdot\lambda_{\min}$ is correct up to a deviation of $2\cdot\max\{\frac{\ln(\frac{1}{\theta-3/20})}{\ln\frac{1}{\theta}}, \frac{\ln(\frac{1}{\theta})}{\ln(\frac{1}{\theta+3/20})}\}$. In particular, if $1/10 \le \theta \le 9/10$ then the quantile scales by at most

$$\max\begin{cases} 2\,\dfrac{\ln(10)}{\ln(4)} \approx 3.322, & \theta = 1/4,\\[2ex] 2\,\dfrac{\ln(\frac{4}{3})}{\ln(\frac{10}{9})} \approx 5.461, & \theta = 9/10. \end{cases}$$

then a clean uniform bound is ≤ 6. Thus, with probability at least $1-\beta$, the output $\tilde{Q}_{1-\theta}$ satisfies

$$\frac{1}{6}q_{1-\theta} \;\le\; \tilde{Q}_{1-\theta} \;\le\; 6\,q_{1-\theta},$$

i.e., Algorithm 1 achieves a constant 6-approximation to the $(1-\theta)$-quantile of P, provided

$$n \ge \max\left\{\frac{5}{\epsilon}\ln\left(\frac{4\log\left(\lambda_{\max}/\lambda_{\min}\right)}{\beta}\right), 200\ln\left(4/\beta\right)\right\}$$

□

Corollary A.5 (Theorem 3.3 restated)**.** Let $\tilde{Q}_{1-\theta}$ be the output of Algorithm 1. Fix a target failure probability $2\beta \in (0,1)$. There exists an absolute constant $c_0 \ge 1$ from Theorem 3.2 such that, if we set

$$\tilde{R} \;=\; C\,\tilde{Q}_{1-\theta}\ln n \quad\text{with}\quad C \;\ge\; \frac{c_0}{\ln(1/\theta)}\Big(1+\frac{\ln(1/\beta)}{\ln n}\Big),$$

then with probability at least $1-2\beta$ all n samples lie in $[0,\tilde{R}]$. In particular,

$$\tilde{R} \;=\; O\left(\frac{1}{\lambda}\ln\Big(1/\theta\Big)\,\ln n\right)$$

Proof. Let $Q_{1-\theta} = \frac{1}{\lambda}\ln(1/\theta)$ be the $(1-\theta)$-quantile of $\mathrm{Exp}(\lambda)$. By Theorem 3.2, there exists c_0 such that with probability at least $1-\beta$:

$$\frac{1}{c_0}\,Q_{1-\theta} \;\le\; \tilde{Q}_{1-\theta} \;\le\; c_0\,Q_{1-\theta}$$

Denote this good event as $\mathcal{E}$ and condition on it. For any sample $X \sim \text{Exp}(\lambda)$,

$$\Pr[X > \tilde{R}] \;\le\; e^{-\lambda \tilde{R}} \;\le\; \exp\Big(-\lambda C \ln n \cdot \tfrac{Q_{1-\theta}}{c_0}\Big) \;=\; \exp\Big(-\tfrac{C}{c_0} \ln n \cdot \ln \tfrac{1}{\theta}\Big) \;=\; n^{-\frac{C}{c_0}\ln(1/\theta)}$$

A union bound over the n samples gives

$$\Pr\big[\max_i X_i > \tilde{R} \;\big|\; \mathcal{E}\big] \;\le\; n^{1-\frac{C}{c_0}\ln(1/\theta)}$$

Choosing $C \ge \frac{c_0}{\ln(1/\theta)}\big(1 + \frac{\ln(1/\beta)}{\ln n}\big)$ ensures the right-hand side is at most β. Unconditioning, the overall failure probability is at most 2β.

Finally, under $\mathcal{E}$:

$$\tilde{R} \;=\; C\,\tilde{Q}_{1-\theta}\,\ln n \;\le\; C\,c_0\,Q_{1-\theta}\,\ln n \;=\; O\Big(\frac{1}{\lambda}\ln\big(1/\theta\big)\,\ln n\Big)$$

which proves the claimed order bound. □

Lemma A.6 (Theorem 3.4 restated)**.** Algorithm 2 is ϵ-differentially private.

Proof. Let $f(P) = \frac{1}{n}\sum_{i=1}^{n} \min\{x_i, R\}$. For neighboring datasets P, P' differing in one record, the ℓ_1-sensitivity satisfies

$$\Delta f \;\le\; \frac{1}{n}\big|\min\{x, R\} - \min\{x', R\}\big| \;\le\; \frac{R}{n}.$$

Adding $\text{Lap}(R/(\epsilon n))$ to $f(P)$ therefore yields an ϵ-DP release by the Laplace mechanism. The final estimator $\tilde{\lambda} = 1/\tilde{s}$ is a deterministic function (post-processing) of the noisy mean $\tilde{s}$, so the overall procedure is ϵ-DP. □

Lemma A.7 (Theorem 3.5 restated)**.** Algorithm 2 returns an $(1 \pm \alpha)$-approximation of λ w.p. $\ge 1 - \beta$, given that:

$$n \ge O\left(\max\left\{\frac{R\ln(\frac{1}{\beta})}{\epsilon(\alpha - e^{-\lambda R})}, \frac{\ln(\frac{1}{\beta})}{\alpha^2}\right\}\right)$$

Proof. Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Exp}(\lambda)$. The maximum likelihood estimate (MLE) of the rate parameter λ is given by:

$$\hat{\lambda} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^{-1}$$

Let $\bar{s} = \frac{1}{n}\sum_{i=1}^{n} \bar{x}_i$ denote the mean of the clipped values, and $\tilde{s}$ the privatized version:

$$\tilde{s} = \bar{s} + \text{Lap}\left(\frac{R}{\epsilon n}\right)$$

Then, the private estimator is:

$$\tilde{\lambda} = \frac{1}{\tilde{s}}$$

We now analyze the error in $\tilde{\lambda}$. First, we bound the deviation of $\tilde{s}$ from the true mean $s = 1/\hat{\lambda}$. The total error in $\tilde{s}$ arises from two sources:

1. **Clipping Bias:** Define $\delta_{\text{clip}} = |\mathbb{E}[\bar{x}_i] - \mu|$. Since $X_i \sim \text{Exp}(\lambda)$, the clipping bias satisfies:

$$\delta_{\text{clip}} = \int_R^{\infty} (x - R)\lambda e^{-\lambda x} dx = e^{-\lambda R}\left(\frac{1}{\lambda}\right)$$

2. **Laplace Noise:** With probability at least $1 - \beta/2$, the Laplace noise is bounded by:

$$\left|\text{Lap}\left(\frac{R}{\epsilon n}\right)\right| \le \frac{R}{\epsilon n}\ln\left(\frac{2}{\beta}\right)$$

Therefore, the total deviation of $\tilde{s}$ from μ is:

$$|\tilde{s} - s| \leq \delta_{\text{clip}} + \left|\text{Lap}\left(\frac{R}{\epsilon n}\right)\right| \leq \frac{e^{-\lambda R}}{\lambda} + \frac{R}{\epsilon n}\ln\left(\frac{2}{\beta}\right)$$

We now analyze the error in $\hat{\lambda}$. We apply a multiplicative Chernoff bound for the sample mean of i.i.d. exponential variables. For any $\alpha \in (0,1)$, we have:

$$\Pr\left[\left|s - \frac{1}{\lambda}\right| \geq \frac{\alpha}{\lambda}\right] \leq 2\exp\left(-\frac{n\alpha^2}{3}\right)$$

Equivalently, since $\hat{\lambda} = 1/s$, this implies:

$$\Pr\left[\left|\frac{\hat{\lambda}}{\lambda} - 1\right| \geq \alpha\right] \leq 2\exp\left(-\frac{n\alpha^2}{3}\right)$$

To ensure this probability is at least $1 - \beta/2$, it suffices to set:

$$2\exp\left(-\frac{n\alpha^2}{3}\right) \leq \beta/2 \implies n \geq \frac{3}{\alpha^2}\ln\left(\frac{4}{\beta}\right)$$

Finally, to guarantee an $(1 \pm \alpha)$-multiplicative approximation to λ, it suffices to require:

$$\left|\tilde{s} - \frac{1}{\lambda}\right| \leq |\tilde{s} - s| + |s - \frac{1}{\lambda}| \leq \frac{\alpha}{\lambda}$$

Thus, we set:

$$\frac{e^{-\lambda R}}{\lambda} + \frac{R}{\epsilon n}\ln\left(\frac{2}{\beta}\right) \leq \frac{\alpha}{2\lambda}$$

This gives the sample complexity condition:

$$n \geq O\left(\max\left\{\frac{R\ln(\frac{1}{\beta})}{\epsilon(\alpha - e^{-\lambda R})}, \frac{\ln(\frac{1}{\beta})}{\alpha^2}\right\}\right)$$

under the assumption that $\alpha > e^{-\lambda R}$. Note that since $R = O\big(\frac{1}{\lambda}\ln(\frac{1}{\theta})\ln(n)\big)$ (see Theorem 3.3), this condition simplifies to $\alpha > 1/n$, which is a mild and natural requirement as typically $n > 1/\alpha$.

Hence, under this condition, the algorithm returns an $(1 \pm \alpha)$-approximation to λ with probability at least $1 - \beta$. □

Lemma A.8 (Theorem 3.7 restated)**.** Algorithm 4 is ϵ-differentially private.

Proof. Let $T = \left\lceil \log_2\left(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\alpha}\right)\right\rceil$ be the maximal number of (binary-search) iterations. Each iteration computes one counting query of sensitivity $1/n$ and releases it with independent Laplace noise of scale $b = T/(\epsilon n)$. By the Laplace mechanism, each such release is (ϵ/T)-DP; by sequential composition over at most T iterations, the full vector of noisy comparisons is ϵ-DP. The final output is a deterministic function of these noisy values, hence by post-processing the algorithm is ϵ-DP. □

Lemma A.9 (Theorem 3.8 restated)**.** Fix target accuracy $\alpha \in (0,1)$ and failure probability $\beta \in (0,1)$. Let $T = \left\lceil \log_2\left(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\alpha}\right)\right\rceil$. If

$$n \;\geq\; \max\left\{\frac{2eT}{\epsilon\alpha}\ln\left(\frac{2T}{\beta}\right), \frac{2}{\alpha^2}\ln\left(\frac{2T}{\beta}\right)\right\}$$

then with probability at least $1 - \beta$ Algorithm 4 outputs $\tilde{\lambda}$ satisfying

$$\tilde{\lambda} \in [(1-\alpha)\lambda,\ (1+\alpha)\lambda]$$

Proof. The binary search compares $\tilde{p}$ with $1-1/e$ at each step. Two error sources affect $\tilde{p}$:

1. **Sampling error.** By the DKW inequality,

$$\Pr\left[\sup_x |F_n(x) - F(x)| > \frac{\alpha}{2}\right] \le 2e^{-2n(\alpha/2)^2}$$

Setting this $\le \beta/(2T)$ and union-bounding over T steps gives the sampling term

$$n \ge \frac{2}{\alpha^2} \ln\left(\frac{2T}{\beta}\right)$$

2. **Noise error.** Recall that the objective of the quantile estimation step is to output a value within the interval $\left[\frac{1}{(1+\alpha)\lambda}, \frac{1}{(1-\alpha)\lambda}\right]$. It suffices to estimate the $1/e$-quantile to within an additive error of $\frac{\alpha}{2e}$ in the CDF domain. Indeed, consider the upper deviation:

$$\int_{1/\lambda}^{1/((1-\alpha)\lambda)} \lambda e^{-\lambda x}\, dx = e^{-1} - e^{-1/(1-\alpha)} = e^{-1}\left(1 - e^{-\frac{\alpha}{1-\alpha}}\right)$$

Using the inequality $e^{-t} \le 1 - t/2$ for $t \in [0,1]$ and noting that $\alpha < 1/2$, we obtain

$$e^{-1}\left(1 - e^{-\frac{\alpha}{1-\alpha}}\right) \ge e^{-1}\left(1 - \left(1 - \frac{\alpha}{2(1-\alpha)}\right)\right) = \frac{\alpha}{2e(1-\alpha)} \ge \frac{\alpha}{2e}$$

Similarly, for the lower deviation we have

$$\int_{1/((1+\alpha)\lambda)}^{1/\lambda} \lambda e^{-\lambda x}\, dx = e^{-1/(1+\alpha)} - e^{-1} = e^{-1}\left(e^{\frac{\alpha}{1+\alpha}} - 1\right)$$

Applying the bound $e^t \ge 1 + t$ for $t \in [0,1]$, we obtain

$$e^{-1}\left(e^{\frac{\alpha}{1+\alpha}} - 1\right) \ge e^{-1}\left(\frac{\alpha}{1+\alpha}\right) \ge \frac{\alpha}{2e}$$

Therefore, an additive $\frac{\alpha}{2e}$-accuracy in the estimated CDF value is sufficient to guarantee that the resulting quantile lies in the desired multiplicative interval for $1/\lambda$.

Laplace noise with scale $b = T/(\epsilon n)$ satisfies $\Pr[|\mathrm{Lap}(b)| > \alpha/2e] \le e^{-(\alpha/2e)/b}$. Setting this $\le \beta/(2T)$ yields

$$n \ge \frac{2eT}{\epsilon\alpha} \ln\left(\frac{2T}{\beta}\right)$$

When both bounds hold, at each step the comparison with $1-1/e$ is correct, ensuring that after T steps the returned $\tilde{q}$ lies in

$$\left[\frac{1}{(1+\alpha)\lambda}, \frac{1}{(1-\alpha)\lambda}\right]$$

Inverting yields $\tilde{\lambda} \in [(1-\alpha)\lambda, (1+\alpha)\lambda]$. □

Lemma A.10 (Theorem 3.10 restated)**.** Algorithm 5 is ϵ-differentially private.

Proof. The algorithm proceeds in two stages: (i) it runs the quantile-based estimator with budget $\epsilon/3$ and obtains $\tilde{\lambda}_0$; (ii) conditioned on this *released* value, it runs either the MLE route or the quantile route with fresh budget $2\epsilon/3$. By the post-processing property, choosing the branch based on $\tilde{\lambda}_0$ does not incur additional privacy loss, and by adaptive sequential composition the total privacy loss is at most $\epsilon/3 + 2\epsilon/3 = \epsilon$. □

Lemma A.11 (Theorem 6.1 restated)**.** Algorithm 6 is (ϵ, δ)-differentially private. Moreover, if

$$n \geq \max\left\{\frac{800}{\epsilon}\ln(\frac{2}{\delta\beta}), 5000\ln(\frac{2}{\beta})\right\}$$

then

$$\Pr[2^{k^*-1} < {}^{\ln(2)}/_{\lambda} < 2^{k^*+1}] \geq 1-\beta$$

Proof. A proof of (ϵ, δ)-differential privacy of this algorithm can be found in Theorem 3.5 of Vadhan (2017). For utility, we show that if

$$n \geq \max\left\{\frac{800}{\epsilon}\ln(\frac{2}{\delta\beta}), 5000\ln(\frac{2}{\beta})\right\}$$

then $\Pr[2^{k^*-1} < {}^{\ln(2)}/_{\lambda} < 2^{k^*+1}] \geq 1-\beta$. Let $m = {}^{\ln(2)}/_{\lambda}$ be the population median and let $p_k = \Pr[X_i \in (2^k, 2^{k+1})] = e^{-\lambda 2^k} - e^{-\lambda 2^{k+1}}$. The map $k \mapsto p_k$ is unimodal and maximized at indices k with $\lambda 2^k$ closest to $\ln(2)$, hence the top two masses are at $k_m = \lfloor \log_2 m \rfloor$ and $k_m + 1$, with $p_k \in [0.207, 0.25]$ by a direct calculation. Lemma 2.3 of Karwa and Vadhan (2017) gives that if $n \geq \max\{\frac{800}{\epsilon}\ln(\frac{2}{\delta\beta}), 5000\ln(\frac{2}{\beta})\}$ then $\Pr[|\tilde{p}_{k_m} - p_{k_m}| > {}^1/_{100}] \leq \beta$. Laplace noise with scale $2/(\varepsilon n)$ and the threshold $\frac{2}{\epsilon n}\ln({}^2/_{\delta}) + \frac{1}{n}$ ensure that (i) no empty bin enters S with probability $1-\delta$, and (ii) the perturbation cannot change the winner outside $\{k_m, k_m+1\}$. Therefore $k^* \in \{k_m, k_m+1\}$ with probability at least $1-\beta$, and the expanded interval $(2^{k^*-1}, 2^{k^*+1})$ contains $(2^{k_m}, 2^{k_m+1}) \ni m$. Rearranging we have:

$$\begin{aligned} 2^{k^*-1} < {}^{\ln(2)}/_{\lambda} < 2^{k^*+1} &\iff (2^{k^*+1})^{-1} < {}^{\lambda}/_{\ln(2)} < (2^{k^*-1})^{-1} \\ &\iff \ln(2)(2^{k^*+1})^{-1} < \lambda < \ln(2)(2^{k^*-1})^{-1} \iff \tilde{\lambda}_{\min} < \lambda < \tilde{\lambda}_{\max} \end{aligned}$$

□

Lemma A.12 (Theorem 5.1 restated)**.** Let $T(r)$ denote the total variation distance between two exponential distributions whose rates differ by a ratio $r \geq 1$:

$$\begin{aligned} T(r) &:= \mathrm{TV}\big(\mathrm{Exp}(r\lambda), \mathrm{Exp}(\lambda)\big) = r^{-\frac{1}{r-1}} - r^{-\frac{r}{r-1}} \\ &= r^{-\frac{1}{r-1}}\big(1 - \tfrac{1}{r}\big) \end{aligned}$$

Fix any $\alpha \in (0, \frac{1}{2})$. Then there exists an absolute constant c such that, with $r = 1 + c\alpha$, one has $T(r) \geq \alpha$. In particular, the choice $c = 8$ suffices.

Proof. Write $r = 1 + \delta$ with $\delta > 0$ and define

$$f(\delta) := T(1+\delta) = \frac{\delta}{1+\delta}\,(1+\delta)^{-1/\delta},$$

$$R(\delta) := \frac{f(\delta)}{\delta} = \frac{(1+\delta)^{-1/\delta}}{1+\delta}$$

We prove that R is strictly decreasing on $(0, \infty)$. Indeed,

$$\ln R(\delta) \;=\; -\frac{\ln(1+\delta)}{\delta} - \ln(1+\delta)$$

so

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\delta}\ln R(\delta) &= -\left(\frac{1}{\delta(1+\delta)} - \frac{\ln(1+\delta)}{\delta^2}\right) - \frac{1}{1+\delta} \\ &= \frac{(1+\delta)\ln(1+\delta) - \delta - \delta^2}{\delta^2(1+\delta)} \end{aligned}$$

Let $\psi(\delta) := (1+\delta)\ln(1+\delta) - \delta - \delta^2$. Then

$$\psi'(\delta) = \ln(1+\delta) - 2\delta < \delta - 2\delta = -\delta < 0 \quad (\delta > 0)$$

because $\ln(1+x) < x$ for $x > 0$. Since $\psi(0) = 0$ and ψ' is negative on $(0, \infty)$, we have $\psi(\delta) < 0$ for all $\delta > 0$. Hence $(\ln R)'(\delta) < 0$, so R is strictly decreasing.

Consequently, for every $\delta \in (0, 4]$,

$$R(\delta) \;\geq\; R(4) = \frac{(1+4)^{-1/4}}{1+4} = \frac{1}{5^{5/4}}$$

Now fix $c = 8$ and take any $\alpha \in (0, \frac{1}{2})$. Then $\delta = c\alpha \in (0, 4)$ and

$$T(1+c\alpha) \;=\; f(\delta) \;=\; \delta\, R(\delta) \;\geq\; \delta\, R(4) \;=\; \frac{c}{5^{5/4}}\,\alpha$$

Since $8 \geq 5^{5/4}$, we obtain $T(1+8\alpha) \geq \alpha$ for all $\alpha \in (0, \frac{1}{2})$, as required. □

B Applications to Pareto Distributions

The exponential distribution is not only fundamental in its own right but also serves as a key building block for other families. In particular, the *Pareto distribution*—one of the most widely studied heavy-tailed laws in economics, finance, and networks—admits a simple logarithmic reduction to the exponential. This reduction allows us to transfer our private exponential learning algorithms directly to the Pareto setting.

B.1 Learning a Pareto Distribution

The techniques developed for exponential distribution can also be adapted to privately learn the parameters of a *Pareto distribution*, which models heavy-tailed phenomena. Recall that the standard Pareto distribution, that is Pareto(x_m, α_p), is defined over $x \geq x_m > 0$ with probability density function:

$$f(x) = \alpha_p \cdot \frac{x_m^{\alpha_p}}{x^{\alpha_p+1}} \qquad \text{for } x \geq x_m,$$

where x_m is the minimum value and $\alpha_p > 0$ is the shape parameter.

We consider the case where x_m is *known and fixed*, and the goal is to estimate the shape parameter α_p. The MLE estimator is:

$$\hat{\alpha_p} = \left(\frac{1}{n} \sum_{i=1}^{n} \ln\left(\frac{x_i}{x_m}\right) \right)^{-1}$$

This estimator closely resembles the exponential MLE, since $\ln(x/x_m) \sim \text{Exp}(\alpha_p)$ when $x \sim \text{Pareto}(x_m, \alpha_p)$. This logarithmic transformation allows us to reduce the problem of learning a Pareto distribution to that of learning an exponential distribution, which we already handle.

To learn α_p under differential privacy:

1. **Log-transform the data:** For each $x_i \in P$, define $y_i = \ln(x_i/x_m)$. Then $y_i \sim \text{Exp}(\alpha_p)$.
2. **Apply Algorithm 3:** Use our private exponential learning algorithm on the transformed dataset $\{y_i\}_{i=1}^n$ to obtain a private estimate $\tilde{\alpha}_p$.

Lemma B.1. Let $P(x_m, \alpha_p)$ denote the Pareto distribution. For two parameter pairs (x_m, α_p) and $(\tilde{x}_m, \tilde{\alpha}_p)$ set

$$\Delta_s := \left|\ln(\tilde{x}_m/x_m)\right|, \qquad \alpha_{\min} := \min\{\alpha_p, \tilde{\alpha}_p\}, \quad \alpha_{\max} := \max\{\alpha_p, \tilde{\alpha}_p\}$$

Then the total variation distance satisfies

$$\text{TV}(P(x_m, \alpha_p),\, P(\tilde{x}_m, \tilde{\alpha}_p)) \;\leq\; 1 - e^{-\alpha_{\max}\Delta_s} \;+\; \sqrt{\tfrac{1}{2}\left(\tfrac{\alpha_{\max}}{\alpha_{\min}} - 1 - \ln\tfrac{\alpha_{\max}}{\alpha_{\min}}\right)}$$

and, for small gaps, the linearized bound TV $\leq\; \alpha_{\max}\Delta_s + \frac{|\tilde{\alpha}_p - \alpha_p|}{2\alpha_{\min}}$ also holds.

Proof. By the triangle inequality in total variation,

$$\mathrm{TV}\big(P(x_m,\alpha_p),P(\tilde{x}_m,\tilde{\alpha}_p)\big) \le \mathrm{TV}\big(P(x_m,\alpha_p),P(\tilde{x}_m,\alpha_p)\big) + \mathrm{TV}\big(P(\tilde{x}_m,\alpha_p),P(\tilde{x}_m,\tilde{\alpha}_p)\big)$$

The first term compares equal-tail-index Paretos and depends only on scale. For a fixed tail index $\bar{\alpha}$, the exact formula is $\mathrm{TV}\big(P(x_m,\bar{\alpha}),P(\tilde{x}_m,\bar{\alpha})\big) = 1-\exp(-\bar{\alpha}\,|\ln(\tilde{x}_m/x_m)|)$; taking $\bar{\alpha}=\alpha_{\max}$ gives the stated upper bound $1-e^{-\alpha_{\max}\Delta_s}$.

For the second term (equal scale $\tilde{x}_m$), apply Pinsker's inequality $\mathrm{TV} \le \sqrt{\frac{1}{2}\,\mathrm{KL}}$ together with the closed-form KL between equal-scale Paretos:

$$\mathrm{KL}(P(\tilde{x}_m,\alpha_p)\,\|\,P(\tilde{x}_m,\tilde{\alpha}_p)) = \frac{\alpha_p}{\tilde{\alpha}_p} - 1 - \ln\frac{\alpha_p}{\tilde{\alpha}_p}$$

By symmetry in $(\alpha_p,\tilde{\alpha}_p)$ this equals $\frac{\alpha_{\max}}{\alpha_{\min}} - 1 - \ln\frac{\alpha_{\max}}{\alpha_{\min}}$, and the stated bound follows. The linearized bound is the first-order expansion $1-e^{-t}\le t$ and $u-1-\ln u \le \frac{1}{2}(u-1)^2$ near $t=0$, $u=1$. □

Corollary B.2. Let $P(x_m,\alpha_p)$ and $P(\tilde{x}_m,\tilde{\alpha}_p)$ be two Pareto distributions. If

$$\tilde{\alpha}_p \in \big[(1-\gamma)\alpha_p,\ (1+\gamma)\alpha_p\big] \quad \text{and} \quad \frac{\tilde{x}_m}{x_m} \le e^{\frac{\gamma}{2\alpha_p}}$$

then $\mathrm{TV}(P(x_m,\alpha_p),\,P(\tilde{x}_m,\tilde{\alpha}_p)) \ \le\ \gamma$.

The reduction from Pareto to exponential distributions allows us to directly transfer our best-of-both exponential learner into the Pareto setting. By applying the logarithmic transform to the input samples, running the exponential algorithm, and then inverting the transform, we obtain an efficient private learner for Pareto distributions. The following corollary summarizes the resulting guarantees.

Corollary B.3. Given n samples from Pareto(x_m,α_p), the composition of the log-transform with Algorithm 3 yields an ϵ-differentially private estimate $\tilde{\alpha}_p$ such that, with probability at least $1-\beta$,

$$|\tilde{\alpha}_p - \alpha_p| \le \gamma\alpha_p$$

for any $\gamma\in(0,1)$, provided that

$$n \ge O\left(\max\left\{\min\left\{\frac{\ln(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\alpha})}{\epsilon\gamma}\ln(\frac{\ln(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\alpha})}{\beta}),\ \frac{\ln(\frac{1}{\alpha})\ln(\frac{1}{\beta})}{\alpha_p\epsilon\gamma}\right\},\ \frac{\ln(\frac{\ln(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\alpha})}{\beta})}{\gamma^2},\ \frac{\ln(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\beta})}{\epsilon}\right\}\right)$$

which yields $\mathrm{TV}(P(x_m,\alpha_p),\,P(x_m,\tilde{\alpha}_p)) \ \le\ \gamma$ by Theorem B.2.

B.1.1 Estimating the Scale Parameter x_m Under Differential Privacy

We now address the case where both the shape α_p and the scale x_m of a Pareto distribution Pareto(x_m,α_p) are unknown. A direct private release of the sample minimum is problematic due to its large global sensitivity. Instead, we leverage a *quantile-tail* reduction that preserves both privacy and statistical efficiency.

Key observation (scale invariance). If $X\sim\mathrm{Pareto}(x_m,\alpha_p)$ and $t\ge x_m$, then conditional on $X\ge t$ we have the exact scaling law

$$\ln\left(\frac{X}{t}\right)\ \Big|\ \{X\ge t\}\ \sim\ \mathrm{Exp}(\alpha_p)$$

Lemma B.4 (Tail-log transform of Pareto is exponential)**.** Let $X\sim\mathrm{Pareto}(x_m,\alpha_p)$ with CDF $F(x)=1-(x_m/x)^{\alpha_p}$ for $x\ge x_m$, and fix any $t\ge x_m$. Then

$$\ln\left(\frac{X}{t}\right)\ \Big|\ \{X\ge t\}\ \sim\ \mathrm{Exp}(\alpha_p)$$

Proof. For $y \geq 0$,

$$\Pr\left[\ln\left(\frac{X}{t}\right) > y \;\middle|\; X \geq t\right] \;=\; \frac{\Pr\left[\ln(X/t) > y,\; X \geq t\right]}{\Pr[X \geq t]} \;=\; \frac{\Pr\left[X > te^y\right]}{\Pr[X \geq t]}$$

Since X is Pareto,

$$\Pr[X > u] = \left(\frac{x_m}{u}\right)^{\alpha_p} \qquad \text{for } u \geq x_m.$$

Applying this with $u = te^y$ and with $u = t$ gives

$$\Pr\left[\ln\left(\frac{X}{t}\right) > y \;\middle|\; X \geq t\right] \;=\; \frac{(x_m/(te^y))^{\alpha_p}}{(x_m/t)^{\alpha_p}} \;=\; e^{-\alpha_p y}$$

Thus the conditional tail function is $e^{-\alpha_p y}$ for $y \geq 0$, which is exactly the tail of $\mathrm{Exp}(\alpha_p)$. Hence $\ln(X/t) \mid \{X \geq t\} \sim \mathrm{Exp}(\alpha_p)$ □

Thus, if we can privately estimate a low quantile q_τ (for a small fixed $\tau \in (0,1)$) and then look only at exceedances above q_τ, the log-exceedances behave as i.i.d. exponentials with rate α_p, independently of the unknown x_m.

Lemma B.5 (Pareto Quantile and Scale Identity)**.** Let $X \sim \mathrm{Pareto}(x_m, \alpha_p)$ with CDF $F(x) = 1 - (x_m/x)^{\alpha_p}$ for $x \geq x_m$, where $x_m > 0$ and $\alpha_p > 0$. For any $\tau \in (0,1)$, the τ-quantile q_τ satisfies

$$q_\tau \;=\; \frac{x_m}{(1-\tau)^{1/\alpha_p}}.$$

Equivalently,

$$x_m \;=\; q_\tau\,(1-\tau)^{1/\alpha_p}.$$

Proof. By definition of the τ-quantile, q_τ is any value such that $F(q_\tau) = \tau$. For $X \sim \mathrm{Pareto}(x_m, \alpha_p)$ and $q_\tau \geq x_m$,

$$F(q_\tau) \;=\; 1 - \left(\frac{x_m}{q_\tau}\right)^{\alpha_p} \;=\; \tau \;\Longrightarrow\; \left(\frac{x_m}{q_\tau}\right)^{\alpha_p} \;=\; 1-\tau.$$

Taking $1/\alpha_p$ powers (noting $\alpha_p > 0$) yields

$$\frac{x_m}{q_\tau} \;=\; (1-\tau)^{1/\alpha_p} \;\Longrightarrow\; q_\tau \;=\; \frac{x_m}{(1-\tau)^{1/\alpha_p}}.$$

Rearranging gives the equivalent identity $x_m = q_\tau(1-\tau)^{1/\alpha_p}$. □

Remark. Lemma B.5 is just the closed-form quantile function $F^{-1}(\tau)$ of the Pareto distribution. It underlies our scale recovery step: given (private) estimates $(\hat{q}_\tau, \tilde{\alpha}_p)$, we set $\tilde{x}_m = \hat{q}_\tau(1-\tau)^{1/\tilde{\alpha}_p}$.

Having established the logarithmic reduction from Pareto to exponential distributions, we can now lift our adaptive best-of-both exponential learner directly to the Pareto setting. The idea is simple: first apply the log-transform to convert Pareto samples into exponential ones, then run the best-of-both exponential learner under differential privacy, and finally invert the transformation to recover a private estimate of the Pareto parameter. The complete procedure is summarized in Algorithm 7.

Algorithm 7 DP Learning of α_p and x_m for $\mathrm{Pareto}(x_m, \alpha_p)$

Input: $P = \{x_i\}_{i=1}^n$, bounds $0 < \lambda_{\min} < \lambda_{\max}$, target error $\alpha \in (0,1)$, privacy budget $\epsilon > 0$.
Output: $(\tilde{\alpha}_p, \tilde{x}_m)$.

$\tau \leftarrow \frac{1}{4\ln(7)}$
$\hat{q}_\tau \leftarrow$ QUANTILE-ESTIMATION$(P,\ \lambda_{\min},\ \lambda_{\max},\ \theta = 1-\tau,\ \epsilon/2)$ ▷ DP low-quantile
Let $I = \{i : x_i \geq \hat{q}_\tau\}$ and define $y_i = \ln(x_i/\hat{q}_\tau)$ for $i \in I$. ▷ Tail transform
$\tilde{\alpha}_p \leftarrow$ BEST-OF-BOTH$(\{y_i\}_{i\in I},\ \lambda_{\min} = \hat{q}_\tau,\ \lambda_{\max},\ \alpha,\ \epsilon/2)$
Scale recovery:

$$\tilde{x}_m \;\leftarrow\; \hat{q}_\tau \cdot (1-\tau)^{1/\tilde{\alpha}_p} \qquad \text{since} \qquad q_\tau \;=\; \frac{x_m}{(1-\tau)^{1/\alpha_p}}.$$

return $(\tilde{\alpha}_p, \tilde{x}_m)$

Lemma B.6 (Privacy of Algorithm 7)**.** Algorithm 7 is ϵ-differentially private.

Proof. The algorithm first runs the DP low-quantile routine with budget $\epsilon/2$; this is $\epsilon/2$-DP. It then forms the (deterministic) transformed dataset $\{y_i = \ln(x_i/\hat{q}_\tau) : x_i \geq \hat{q}_\tau\}$ using the already released $\hat{q}_\tau$ and the raw data, and applies a second DP estimator (the BEST-OF-BOTH exponential learner) with budget $\epsilon/2$ to $\{y_i\}$. For any fixed (public) $\hat{q}_\tau$, this second mechanism is $\epsilon/2$-DP with respect to the underlying dataset. Hence, by adaptive sequential composition the pair of releases is ϵ-DP. The final scale estimate $\tilde{x}_m = \hat{q}_\tau(1-\tau)^{1/\tilde{\alpha}_p}$ is a deterministic function of the two released values, so by post-processing the overall algorithm is ϵ-DP. □

Lemma B.7 (Utility of Algorithm 7)**.** Fix $\tau \in (0, 1/4]$ and target error $\gamma \in (0, 1/2]$ and $\beta \in (0, 1)$. If

$$n \;\geq\; \frac{1}{1-\tau}\cdot O\left(\max\left\{\min\left\{\frac{\ln(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\alpha})}{\epsilon\gamma}\ln(\frac{\ln(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\alpha})}{\beta}), \frac{\ln(\frac{1}{\alpha})\ln(\frac{1}{\beta})}{\alpha_p\epsilon\gamma}\right\}, \frac{\ln(\frac{\ln(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\alpha})}{\beta})}{\gamma^2}, \frac{\ln(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\beta})}{\epsilon}\right\}\right)$$

then with probability at least $1-\beta$ the outputs of Algorithm 7 satisfy

$$\tilde{\alpha}_p \in \big[(1-\gamma)\alpha_p,\ (1+\gamma)\alpha_p\big] \quad \text{and} \quad \frac{\tilde{x}_m}{x_m} \leq e^{\ 2\ln(7)\,(\gamma/\alpha_p)\tau}$$

for a universal constant $c > 0$. In particular, choosing $\tau = \frac{1}{4\ln(7)}$ yields a sufficient estimate of x_m which yields $\mathrm{TV}(P(x_m, \alpha_p),\, P(\tilde{x}_m, \tilde{\alpha}_p)) \;\leq\; \gamma$ by Theorem B.2.

Proof. Write $n_\tau := |I|$ for the (random) number of exceedances above $\hat{q}_\tau$. In expectation, $n_\tau \approx (1-\tau)n$. The accuracy of $(\tilde{\alpha}_p, \tilde{x}_m)$ follows from three components:

1. **Low-quantile accuracy.** By Theorem 3.2 with $\theta = 1-\tau$ and confidence $\beta/2$, the private estimate $\hat{q}_\tau$ is a multiplicative $O(1)$-approximation to q_τ provided
$$n \geq \max\left\{\frac{20}{\epsilon}\ln\left(\frac{4\log\left(\lambda_{\max}/\lambda_{\min}\right)}{\beta}\right), 200\ln\left(4/\beta\right)\right\}$$

2. **Rate accuracy on exceedances.** Conditioned on $\hat{q}_\tau$, the variables $\{y_i\}_{i\in I}$ are i.i.d. $\mathrm{Exp}(\alpha_p)$. Applying Theorem 3.11 with sample size n_τ and confidence $\beta/2$ yields a multiplicative $(1\pm\gamma)$ estimate $\tilde{\alpha}_p$ when
$$n_\tau \geq C\cdot\max\left\{\min\left\{\frac{\ln(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\alpha})}{\epsilon\gamma}\ln(\frac{\ln(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\alpha})}{\beta}), \frac{\ln(\frac{1}{\alpha})\ln(\frac{1}{\beta})}{\alpha_p\epsilon\gamma}\right\}, \frac{\ln(\frac{\ln(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\alpha})}{\beta})}{\gamma^2}, \frac{\ln(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\beta})}{\epsilon}\right\}$$
for a universal constant C.

3. **Scale recovery error propagation.** Since $x_m = q_\tau\,(1-\tau)^{1/\alpha_p}$ and $\tilde{x}_m = \hat{q}_\tau\,(1-\tau)^{1/\tilde{\alpha}_p}$,
$$\frac{\tilde{x}_m}{x_m} \;=\; \underbrace{\frac{\hat{q}_\tau}{q_\tau}}_{\text{quantile factor}} \cdot \underbrace{(1-\tau)^{\frac{1}{\tilde{\alpha}_p}-\frac{1}{\alpha_p}}}_{\text{rate factor}} \;\leq\; 7\cdot\exp\Big(\Big(\tfrac{1}{\tilde{\alpha}_p}-\tfrac{1}{\alpha_p}\Big)\ln(1-\tau)\Big)$$
If $\tilde{\alpha}_p \in [(1-\gamma)\alpha_p, (1+\gamma)\alpha_p]$, then
$$\Big|\tfrac{1}{\tilde{\alpha}_p}-\tfrac{1}{\alpha_p}\Big| \;\leq\; \frac{\gamma}{(1-\gamma)\,\alpha_p} \;\leq\; \frac{2\gamma}{\alpha_p} \quad \text{for } \gamma \leq \tfrac{1}{2},$$
and hence
$$(1-\tau)^{\frac{1}{\tilde{\alpha}_p}-\frac{1}{\alpha_p}} \leq \exp\Big(\tfrac{2\gamma}{\alpha_p}\,|\ln(1-\tau)|\Big)$$
For $\tau \leq \frac{1}{4}$, $|\ln(1-\tau)| \leq \frac{4}{3}\tau$. Thus,
$$\frac{\tilde{x}_m}{x_m} \in 7\cdot\exp\Big(\tfrac{2\gamma}{\alpha_p}\,\tau\Big)$$

□

We now state the formal guarantees for the private Pareto learner. Since Algorithm 7 relies only on the logarithmic reduction and then invokes the exponential best-of-both learner, its privacy and accuracy properties follow directly from our earlier results. The theorem below shows that the algorithm achieves a $(1 \pm \gamma)$ approximation to the Pareto parameter under differential privacy, which directly yields a γ-bound on the total variation distance between the learned and the true distributions, with essentially the same sample complexity as in the exponential case up to logarithmic factors.

Theorem B.8. *Fix $\tau \in (0, 1/4]$, target accuracy parameter $\gamma \in (0, 1/2]$, failure probability $\beta \in (0, 1)$, and privacy budget $\epsilon > 0$. Then Algorithm 7 is ϵ-differentially private.*

Moreover, if the sample size n satisfies

$$n \;\geq\; \frac{C}{1-\tau}\cdot\max\left\{\min\left\{\frac{\ln(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\alpha})}{\epsilon\gamma}\ln(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\alpha\beta}), \frac{\ln(\frac{1}{\alpha})\ln(\frac{1}{\beta})}{\alpha_p\epsilon\gamma}\right\}, \frac{\ln(\frac{\ln(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\alpha})}{\beta})}{\gamma^2}, \frac{\ln(\frac{\log(\lambda_{\max}/\lambda_{\min})}{\beta})}{\epsilon}\right\}$$

for a universal constant $C > 0$, then with probability at least $1-\beta$ the outputs $(\tilde{\alpha}_p, \tilde{x}_m)$ satisfy

$$\tilde{\alpha}_p \;\in\; \big[(1-\gamma)\alpha_p,\; (1+\gamma)\alpha_p\big]$$

and

$$\frac{\tilde{x}_m}{x_m} \;\leq\; e^{\,c\,(\gamma/\alpha_p)\tau}$$

for a universal constant $c > 0$.

In particular, choosing $\tau = \frac{1}{4\ln(7)}$ yields a sufficient estimate of x_m which yields $\mathrm{TV}(P(x_m, \alpha_p),\, P(\tilde{x}_m, \tilde{\alpha}_p)) \;\leq\; \gamma$ by Theorem B.2.

Proof. Privacy is immediate from Theorem B.6. Utility follows from Theorem B.7. □

This application illustrates how exponential learning algorithms serve as a building block for differentially private estimation in more general parametric families.