UNIFYING PRIVACY ACCOUNTING: INFORMATION EQUIVALENCE AND INFORMATION LOSS

Anonymous authors
Paper under double-blind review

ABSTRACT

Differential privacy (DP) admits several notions, but the choice among them may affect both privacy analysis and utility. In this paper, we consider four mainstream curve-based privacy notions within a unified information-theoretic framework. For a fixed ordered pair of output distributions, we establish information equivalence among the two directional privacy profiles of (ε, δ)-DP, the hypothesis-testing trade-off function, and the extended privacy-loss distribution. The exact Rényi differential privacy (RDP) curve joins this equivalence class whenever it is finite at some order greater than one. Under this mild condition, choosing among these notions changes only their semantic interpretation and computational requirements. In contrast, taking the maximum of the directional privacy profiles or compressing the RDP curve into a single zero-concentrated differential privacy (zCDP) parameter can lose information. We quantify the information loss between the exact RDP curve and its zCDP bound for standard noise mechanisms. This gap is zero for Gaussian noise but positive for Gaussian-mixture, Laplace, discrete Gaussian, and Poisson-subsampled Gaussian mechanisms. Moreover, this gap grows linearly as the number of folds of composition increases. Our information-theoretic perspective has practical consequences. At the same certified privacy level, retaining the full RDP curve rather than using zCDP reduces the required noise variance by up to 45% for Gaussian-mixture noise in workloads comparable in size to the American Community Survey. For DP-SGD on Fashion-MNIST under Poisson subsampling, an RDP-based privacy accountant improves test accuracy by up to 8.73 percentage points compared to a zCDP-based accountant under the same (ε, δ) guarantee obtained by converting from RDP and zCDP. These results clarify when different privacy notions have equivalent information (when the gap is zero or nearly zero) and when retaining more detailed privacy information improves utility (when the gap is large).

1 INTRODUCTION

Differential privacy (DP) is a rigorous mathematical framework for quantifying changes in the output distributions of a randomized algorithm on two neighboring datasets that differ in at most one element (Dwork, 2006; Dwork & Roth, 2014). For a fixed ordered pair of output distributions induced by a mechanism on neighboring datasets, different privacy notions serve different purposes: Rényi differential privacy (RDP) supports convenient composition through moment accounting and is widely used for DP-SGD (Mironov, 2017; Abadi et al., 2016; Wang et al., 2019); f-DP provides a hypothesis-testing trade-off function and a tensor-product composition rule (Dong et al., 2022; Su, 2025; Su et al., 2025); privacy-loss distributions (PLDs) convolve under product composition and support sharp numerical accounting (Koskela et al., 2020; Gopi et al., 2021; Koskela et al., 2021); and zero-concentrated differential privacy (zCDP) compresses the RDP curve into a single parameter ρ, enabling simple scalar composition (Bun & Steinke, 2016). Meanwhile, (ε, δ)-DP remains the standard reporting format, and its optimal composition admits an operational hypothesis-testing interpretation (Kairouz et al., 2017).

However, a single (ε, δ) pair gives only one operating point of the full privacy profile curve. Such pairs also do not induce a total ordering: for example, $(2, 10^{-6})$ and $(1, 10^{-5})$ are generally incomparable. Recent work therefore studies mechanism comparison beyond a single calibration point (Balle et al.,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

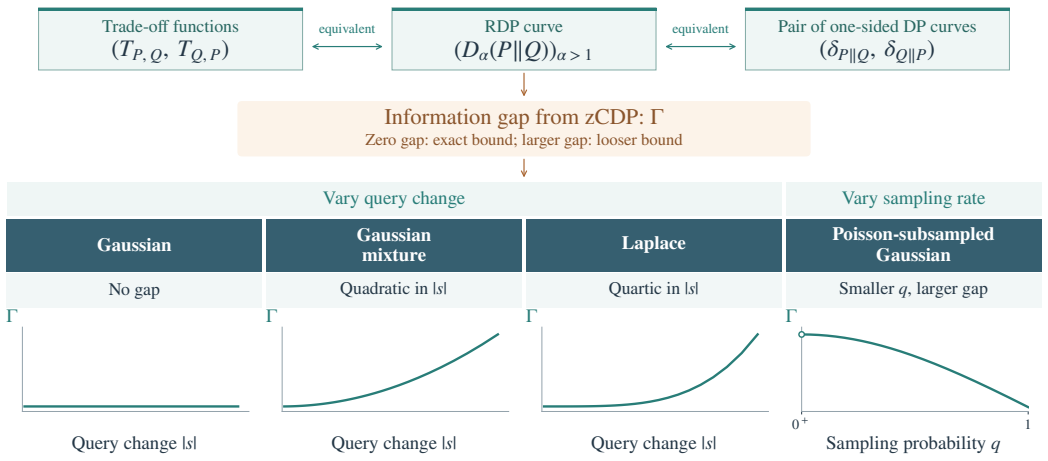


Figure 1: Exact privacy curves and the zCDP–RDP gap. Top: three of the four curve-based notions in Theorem 2.1; the RDP curve joins them under a mild condition. Bottom: how the information gap varies with query change, or sampling probability for four mechanisms. **Left three:** s is the query change between neighboring datasets. At fixed Rényi order $\alpha = 2$, the Gaussian gap is zero, while the symmetric Gaussian-mixture and Laplace gaps scale as $|s|^2$ and $|s|^4$, respectively, as $|s|$ tends to 0. **Right:** q is the rate of a Poisson subsample mechanism. At fixed $\alpha = 2$, smaller $q > 0$ yields a larger gap; $q = 1$ corresponds to no subsampling, and the open circle indicates the limit as $q \to 0$. All panels show trends in $\Gamma(t)$, defined in (3.1). Detailed curves and parameter settings appear in Figure 3.

2020a; Kifer et al., 2022; Kaissis et al., 2024; Gomez et al., 2025). Yet two mechanisms can share the same two-sided (ε, δ) profile, $\max(\delta_{P\|Q}(\varepsilon), \delta_{Q\|P}(\varepsilon))$, while having different exact RDP curves (Section 2). Likewise, Gaussian and mixture noise can share the same optimal zCDP parameter ρ but have different exact RDP curves (Section 3). This leads to two basic questions: (1) Under what conditions do the four full privacy notions, with directional information retained, contain the same privacy information about a fixed pair of output distributions? (2) How much information is lost when a full privacy curve is compressed into a single zCDP parameter?

Answering these questions helps us decide which privacy notion to use. The first question asks whether we can choose a privacy notion to suit our needs without losing privacy information. Prior work establishes duality between complete (ε, δ)-curve guarantees and symmetric f-DP (Dong et al., 2022), as well as optimal conversions from RDP bounds to (ε, δ)-DP and f-DP (Asoodeh et al., 2021; Riess et al., 2026). These conversions are optimal given the RDP bounds, but need not recover the exact privacy curve of a particular distribution pair. In contrast, we ask whether the full exact RDP curve of a fixed ordered pair (P, Q) uniquely determines its directional (ε, δ) profiles, trade-off functions, and PLD. The second question matters in practice: losing privacy information leads to more noise to satisfy the same privacy requirement and reduces accuracy. For example, an f-DP re-accounting of the 2020 U.S. Census (Abowd et al., 2022) showed that the discrete-Gaussian variance proxy σ^2 could be reduced by 15.08%–24.82% while maintaining nearly the same certified privacy guarantees (Su et al., 2025; 2026). In DP-SGD, using the full RDP curve rather than a zCDP summary can also reduce the noise required during privacy calibration, as shown in Figure 2.

Our Contributions. We answer the two questions above through the following contributions, summarized in Figure 1. Related work is discussed in Appendix A.

- **Information equivalence of curve-based privacy notions.** We show that, for a fixed ordered pair (P, Q), the pair of one-sided (ε, δ) profiles, the pair of trade-off functions, and the extended PLD contain the same information, as does the full exact RDP curve under mild conditions. Our counterexamples show that taking the maximum of the one-sided profiles, $\max(\delta_{P\|Q}(\varepsilon), \delta_{Q\|P}(\varepsilon))$, can lose information.

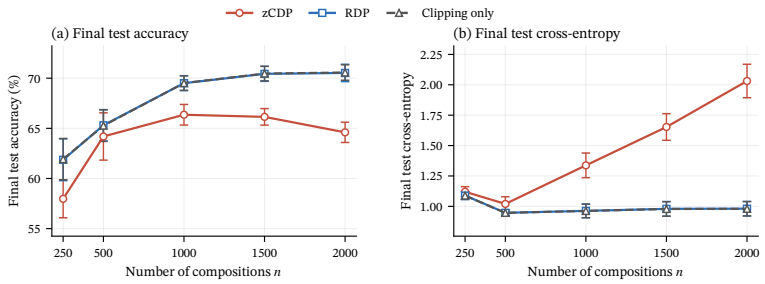


Figure 2: DP-SGD on Fashion-MNIST with noise calibrated under RDP and under zCDP to the same certified privacy $(\varepsilon, \delta) = (3, 10^{-5})$; "clipping only" adds no noise. Final test accuracy and cross-entropy (lower is better) as the number of training steps n varies at sampling rate $q = 0.01$, with the noise recalibrated for each n. Means over five seeds; bars are 95% confidence intervals. Setup in Section 5.2; training trajectories in Appendix E.3.

- **Information loss of zCDP for common mechanisms.** We introduce the zCDP–RDP gap, the distance between the zCDP bound and the exact RDP curve, as a measure of the information discarded when the curve is compressed into a single parameter. The gap adds up under independent composition, and we expand it for small query changes s (Theorem 3.1). It is identically zero for Gaussian noise, of order s^2 for Gaussian-mixture noise, and of order s^4 for Laplace noise. Under Poisson subsampling with rate q it grows as q decreases and approaches the entire zCDP bound as $q \to 0$: zCDP discards the privacy amplification that subsampling provides.
- **Practical guidance for privacy accounting.** Experiments on private data release and DP-SGD show that information loss can require additional noise and reduce accuracy. Our results support using the full RDP curve or a privacy notion with equivalent information when tighter accounting is needed, and zCDP when the zCDP–RDP gap is negligible.

2 Information Equivalence Between Exact Privacy Notions

Fix two neighboring datasets and let P and Q be the output distributions of the mechanism on them. For $\varepsilon \geq 0$, the *one-sided* (ε, δ)*-DP curve* $\delta_{P\|Q}(\varepsilon) := \sup_A \{P(A) - e^{\varepsilon} Q(A)\}$ is the smallest δ such that $P(A) \leq e^{\varepsilon} Q(A) + \delta$ for every event A; exchanging P and Q gives $\delta_{Q\|P}$. The usual *two-sided* curve is $\delta(\varepsilon) = \max\{\delta_{P\|Q}(\varepsilon), \delta_{Q\|P}(\varepsilon)\}$; we keep the two one-sided curves separate. We write $T_{P,Q}$ for the trade-off function, L for the privacy loss, whose distribution under P is the PLD, and $K(t) := \log \mathbb{E}_P[e^{tL}]$ for its cumulant-generating function (CGF). Throughout, $t = \alpha - 1$. At order $\alpha = 1 + t > 1$ the exact RDP curve is $D_{1+t}(P\|Q) = K(t)/t$, and the optimal zCDP parameter is $\rho(P, Q) := \sup_{\alpha > 1} D_\alpha(P\|Q)/\alpha$. Full definitions are in Appendix B.1.

We say that a notion $\mathcal{A}$ *determines* a notion $\mathcal{B}$ if any two pairs (P, Q) and (P', Q') with the same $\mathcal{A}$ also have the same $\mathcal{B}$; two notions that determine each other carry the same information. For instance, the RDP curve determines ρ, but ρ does not determine the RDP curve (Section 3).

Theorem 2.1 (Information equivalence between curve-based notions)**.** *Let P and Q be arbitrary probability measures on $(\Omega, \mathcal{F})$. The following three curve-based notions determine one another:*

(a) *the pair of one-sided (ε, δ)-DP curves $(\delta_{P\|Q}, \delta_{Q\|P})$;*

(b) *the pair of trade-off functions $(T_{P,Q}, T_{Q,P})$;*

(c) *the law of the extended privacy-loss random variable L under P.*

Each notion determines the exact RDP curve $\big(D_\alpha(P\|Q)\big)_{\alpha > 1}$. Conversely, if $D_{\alpha_0}(P\|Q) < \infty$ for some real $\alpha_0 > 1$, the exact RDP curve determines all three curve-based notions above.

Remark. The condition $D_{\alpha_0}(P\|Q) < \infty$ holds for every mechanism in Sections 4 and 5. It fails only when $\mathbb{E}_P[e^{tL}] = \infty$ for every $t > 0$, for example when P puts mass outside the support of Q or

the privacy loss is very heavy-tailed. Although weak, the condition cannot be dropped, nor weakened to a finite mean privacy loss: Example B.2 gives two pairs whose privacy losses have finite means and whose (ε, δ)-DP curves differ, yet whose RDP curves both equal $+\infty$ at every order.

Both one-sided curves are needed. Example B.1 adds one of two integer noise distributions to a count of sensitivity 1. The two distributions have full support, likelihood ratios at most 4, the same right tail, and different left tails; call the resulting pairs (P_A, Q_A) and (P_B, Q_B). For $\varepsilon \geq 0$, the forward curve $\delta_{P\|Q}(\varepsilon)$ involves only outputs where P exceeds Q, here the shared right tail, so both pairs have $\delta_{P\|Q}(\varepsilon) = \frac{2}{11}(4 - e^{\varepsilon})_+$. The reverse curve involves the left tail and differs: $\delta_{Q_A\|P_A}(\log 3) = 3/22$ but $\delta_{Q_B\|P_B}(\log 3) = 0$. Because the reverse curve lies below the forward curve for both pairs, the usual curve $\max\{\delta_{P\|Q}, \delta_{Q\|P}\}$ is the same for both, and the two mechanisms would be reported as equally private. Yet their one-sided curves differ, so by Theorem 2.1 their trade-off functions, PLDs, and RDP curves differ too; indeed $\exp(D_2(P_A\|Q_A)) = 265/88$, whereas $\exp(D_2(P_B\|Q_B)) = 3$. The equivalence in Theorem 2.1 is therefore with the *pair* of one-sided curves; their maximum, the usual curve, carries less information.

When a single curve suffices. Taking the maximum loses no information when the two one-sided curves coincide. This happens for additive noise symmetric about zero, such as Gaussian and Laplace noise: if P and Q are the distributions of $s + Z$ and Z with $Z \overset{d}{=} -Z$, the reflection $x \mapsto s - x$ exchanges P and Q, so $\delta_{P\|Q} = \delta_{Q\|P}$, $T_{P,Q} = T_{Q,P}$, and $D_\alpha(P\|Q) = D_\alpha(Q\|P)$. The usual two-sided curve then carries all the information (Corollary F.1 in Appendix F.2).

Scope of the equivalence. First, Theorem 2.1 concerns information, not computation: converting one notion into another may require the entire curve and need not have a closed form; practical conversion and composition methods are developed in Koskela et al. (2020; 2021); Gopi et al. (2021); Doroshenko et al. (2022); Ghazi et al. (2022). Second, it concerns one fixed pair (P, Q), not a whole mechanism. A mechanism-level RDP guarantee takes, at each order α, the maximum over all neighboring pairs; the maximizing pair may change with α, so the resulting curve need not be the RDP curve of any single pair, and a conversion from it must hold for every pair consistent with it (Zhu et al., 2022; Asoodeh et al., 2021; Balle et al., 2020b). By contrast, replacing the exact RDP curve by the zCDP parameter ρ loses information even for a fixed pair; Section 3 quantifies this loss.

3 THE INFORMATION GAP

By Theorem 2.1, the exact RDP curve, the one-sided curves, the trade-off functions, and the PLD carry the same information. zCDP keeps only ρ, and many curves share the same ρ: for a query change s, Gaussian noise $N(0,1)$ and the mixture $\frac{1}{2}N(-m,1) + \frac{1}{2}N(m,1)$, $m > 0$, both have $\rho = s^2/2$ but different RDP curves (Proposition 4.1). Throughout this section and Section 4, we consider mutually absolutely continuous pairs (P, Q) with finite ρ.

The parameter ρ gives the bound $D_\alpha(P\|Q) \leq \rho\alpha$ at every order $\alpha > 1$; we call the line $\alpha \mapsto \rho\alpha$ the *zCDP envelope*. What zCDP discards is how far the exact curve lies below this line, which we measure by the *information gap* defined as

$$\Gamma(t) := \rho t(t+1) - K(t) = t\left\{(t+1)\rho - D_{t+1}(P\|Q)\right\}, \quad \text{for any } t > 0. \tag{3.1}$$

Here, $(t+1)\rho$ is the zCDP upper bound and $D_{t+1}(P\|Q)$ is the exact RDP value at the same order $\alpha = t + 1$, so $\Gamma(t)$ is their difference multiplied by t. The gap has three useful properties (illustrated in Figure 6, Appendix C). *(1)* $\Gamma(t)/t$ is the vertical distance at order $1 + t$ between the envelope and the exact RDP curve, so $\Gamma \geq 0$, with $\Gamma \equiv 0$ exactly when the envelope reproduces the curve; the factor t puts Γ on the scale of K, which adds up under independent composition. *(2)* For Gaussian noise, $L \sim N(\rho, 2\rho)$ under P, so $K(t) = \rho t(t+1)$ and $\Gamma \equiv 0$: the Gaussian RDP curve *is* its envelope. *(3)* The gap controls the reported ε after order optimization (Proposition 3.2).

3.1 LOCAL EXPANSION

How large is the gap when the noise is large compared with the change one record can cause? For the additive mechanism $\mathcal{M}(D) = \theta(D) + Z$, where Z has density or probability mass function p, let $s := \theta(D) - \theta(D')$ be the query change between neighboring datasets, so $|s|$ is at most the global

sensitivity of θ. Subtracting $\theta(D')$ from both outputs gives the pair (P_s, Q) with densities

$$p_s(x) = p(x - s), \qquad q(x) = p(x). \tag{3.2}$$

More generally, we consider any family (P_s, Q_s) with $P_0 = Q_0$, and study small nonzero s of either sign. For each s, let L_s, K, $\rho(s)$, and Γ be the privacy loss, CGF, optimal zCDP parameter, and gap of (P_s, Q_s), and let $\kappa_j(L_s)$ be the jth cumulant of L_s under P_s (κ_1 is the mean, κ_2 the variance; κ_3 and κ_4 measure skewness and tail weight). Since K and the supremum defining ρ rarely have closed forms, the next theorem reduces Γ for small s to the mean and the third and fourth cumulants of the privacy loss.

Theorem 3.1 (Local expansion of the zCDP–RDP gap)**.** *Fix $T > 0$. Suppose that, for all sufficiently small $s \neq 0$, P_s and Q_s are mutually absolutely continuous, $\rho(s) < \infty$, and the CGF $K(u) = \log \mathbb{E}_{P_s}[e^{uL_s}]$ is five times continuously differentiable on $[-1, T]$ with fifth derivative satisfying $\sup_{-1 \le u \le T} |K^{(5)}(u)| \le M_T |s|^5$, where M_T does not depend on s. Then, for all such s and all $0 < t \le T$, we have*

$$\Gamma(t) = t(t+1)\big(\rho(s) - \mathbb{E}_{P_s} L_s\big) - \frac{t^2(t+1)}{6}\kappa_3(L_s) + \frac{t^2(1-t^2)}{24}\kappa_4(L_s) + \mathcal{R}_s(t), \tag{3.3}$$

with

$$|\mathcal{R}_s(t)| \le \frac{M_T}{120}(t^2 + t^5)|s|^5. \tag{3.4}$$

Remark. The first term measures how far $\rho(s)$ exceeds the mean privacy loss $\mathbb{E}_{P_s} L_s$, which is the KL divergence of P_s from Q_s; the other two involve κ_3 and κ_4. For Gaussian noise all three vanish, since $\rho(s) = \mathbb{E}_{P_s} L_s$ and $\kappa_3 = \kappa_4 = 0$. In general the three terms are of order s^2, s^3, and s^4; Section 4.1 finds the leading one for common mechanisms.

From the gap to privacy bounds. We next relate the information gap to the privacy bounds obtained by applying the same RDP-to-(ε, δ) conversion to the exact RDP curve and its zCDP envelope. Additional more detailed results are given in Appendix C. These comparisons keep the output pair, including its noise scale, unchanged and do not require a small-shift assumption. For a fixed pair (P, Q) satisfying the standing assumptions of this section, let $c_t := t^t/(1+t)^{1+t}$. The conversion of Balle et al. (2020b) implies $\delta_{P\|Q}(\varepsilon) \le c_t \exp\{K(t) - t\varepsilon\}$, with $t > 0$. Its derivation is given in Appendix G.2.

The ε bound at a fixed δ. Fix $\delta \in (0, 1)$ and define $F_{\mathrm{R}}(t) := \frac{K(t) + \log c_t - \log \delta}{t}$, and $F_{\mathrm{Z}}(t) := \rho(t+1) + \frac{\log c_t - \log \delta}{t} = F_{\mathrm{R}}(t) + \frac{\Gamma(t)}{t}$. Minimizing these objectives over $t > 0$ gives the ε upper bounds obtained from the exact RDP curve and its zCDP envelope, respectively. The following proposition quantifies the difference after order optimization.

Proposition 3.2 (Compression loss after order optimization)**.** *Suppose that $\varepsilon_{\mathrm{R}} := \inf_{t>0} F_{\mathrm{R}}(t)$ and $\varepsilon_{\mathrm{Z}} := \inf_{t>0} F_{\mathrm{Z}}(t)$ are attained at $t_{\mathrm{R}}, t_{\mathrm{Z}} \in (0, \infty)$, respectively. Then we have $\frac{\Gamma(t_{\mathrm{Z}})}{t_{\mathrm{Z}}} \le \varepsilon_{\mathrm{Z}} - \varepsilon_{\mathrm{R}} \le \frac{\Gamma(t_{\mathrm{R}})}{t_{\mathrm{R}}}$. Consequently, if an interval $I \subset (0, \infty)$ contains t_{R} and $\sup_{t \in I} \Gamma(t)/t \le \eta$, then $0 \le \varepsilon_{\mathrm{Z}} - \varepsilon_{\mathrm{R}} \le \eta$.*

The relevant quantity is therefore $\Gamma(t)/t$ near the optimizing orders, rather than the gap at an arbitrary reference order. A large value of $\Gamma(t_{\mathrm{Z}})/t_{\mathrm{Z}}$ certifies a nonnegligible compression cost, whereas a uniform upper bound over an interval containing t_{R} certifies that the cost is small. Locating such an interval and bounding the gap without computing the full RDP curve requires mechanism-specific analysis.

Appendix C.1 combines this bound with Theorem 3.1 and gives the analogous δ bound at a fixed ε.

3.2 Information loss under independent composition

We next show how the information gap accumulates under independent composition. For component pairs (P_i, Q_i), let K_i be their privacy-loss CGFs, $\rho_i < \infty$ their optimal zCDP parameters, and $\Gamma_i(t) := \rho_i t(t+1) - K_i(t)$ their information gaps. Define $\boldsymbol{K}(t) := \sum_{i=1}^n K_i(t)$, and $\rho_{\mathrm{sum}} := \sum_{i=1}^n \rho_i$. The gap relative to the standard composed zCDP envelope is $\boldsymbol{\Gamma}(t) := \rho_{\mathrm{sum}} t(t+1) - \boldsymbol{K}(t)$.

Theorem 3.3 (Independent composition)**.** *Let (P_i, Q_i), $i = 1, \dots, n$, be independent component pairs with optimal zCDP parameters $\rho_i < \infty$. Set $P^{(n)} := \bigotimes_{i=1}^n P_i$ and $Q^{(n)} := \bigotimes_{i=1}^n Q_i$. Then, for every $t > 0$, it holds that $tD_{t+1}(P^{(n)} \| Q^{(n)}) = \boldsymbol{K}(t)$ and*

$$\boldsymbol{\Gamma}(t) = \sum_{i=1}^{n} \Gamma_i(t). \tag{3.5}$$

In particular, for n independent copies of a fixed pair (P, Q) with optimal zCDP parameter ρ, the product pair has optimal zCDP parameter $n\rho$ and information gap $n\Gamma(t)$.

Thus, at a fixed order and with the component pair unchanged, the gap grows linearly with the number of independent copies. A component with $P_i = Q_i$ contributes zero.

Standard versus optimal composition. For nonidentical components, the optimal zCDP parameter of the product pair may be smaller than the sum of the component parameters: $\rho_\star := \sup_{t>0} \frac{\boldsymbol{K}(t)}{t(t+1)} \le \rho_{\mathrm{sum}}$. Consequently, we have

$$\boldsymbol{\Gamma}(t) = (\rho_{\mathrm{sum}} - \rho_\star)\, t(t+1) + \big\{\rho_\star t(t+1) - \boldsymbol{K}(t)\big\}. \tag{3.6}$$

Both terms are nonnegative. The first records the difference between adding the component parameters and optimizing the parameter of the product pair; it vanishes for identical components. The second is the gap between the product pair's exact RDP curve and its optimal zCDP envelope.

Under the assumptions of Theorem 3.1 for each component family, the local expansions also add componentwise. Appendix C.2 gives the expansion for heterogeneous shifts and its explicit remainder bound.

4 Information gaps for common mechanisms

4.1 Additive noise mechanisms

We evaluate Γ for four additive noise distributions, Gaussian, Gaussian-mixture, Laplace, and discrete Gaussian, in the setting of (3.2). All four are symmetric about zero, so one direction suffices (Section 2). By Theorem 3.1, the leading power of s in $\Gamma(t)$ measures the loss: $\Gamma \equiv 0$ means none, and the lower the power, the larger the loss for small s.

Proposition 4.1 (Gaps for common additive noise)**.** *Let (P_s, Q) be as in* (3.2) *and fix $T > 0$. The remainders below hold as $s \to 0$, uniformly over $0 < t \le T$.*

(a) Gaussian. *If $Z \sim N(0, \sigma^2)$, then $\rho(s) = s^2/(2\sigma^2)$ and $\Gamma(t) = 0$ for every s and $t > 0$.*

(b) Gaussian mixture. *If $Z \sim \frac{1}{2}N(-m, 1) + \frac{1}{2}N(m, 1)$ with $m \ge 0$, then we have $\rho(s) = s^2/2$ for every s, and*

$$\Gamma(t) = \frac{m^2}{2}\mathbb{E}[\mathrm{sech}^2(mZ)]\, t(t+1)s^2 + O(s^4). \tag{4.1}$$

The coefficient is positive when $m > 0$; for $m = 0$ this is case (a) with $\sigma = 1$.

(c) Laplace. *If Z has density $\frac{1}{2}e^{-|x|}$, then we have $\rho(s) = |s| + e^{-|s|} - 1$ for every s (Harrison & Manurangsi, 2026), and $\Gamma(t) = \frac{[t(t+1)]^2}{12}s^4 + O(|s|^5)$.*

(d) Discrete Gaussian. *If θ is integer-valued, $s = r \in \{1, 2, \dots\}$, and $\mathbb{P}(Z = k) \propto e^{-k^2/(2\sigma^2)}$ on $\mathbb{Z}$ (Canonne et al., 2020), then $\rho = r^2/(2\sigma^2)$ and, with $\Theta_\sigma(u) := \sum_{k\in\mathbb{Z}} e^{-(k-u)^2/(2\sigma^2)}$, $\Gamma(t) = \log \frac{\Theta_\sigma(0)}{\Theta_\sigma((1+t)r)}, t > 0$, which is $1/r$-periodic in t, vanishes exactly when $(1+t)r \in \mathbb{Z}$, and is positive otherwise.*

In short, zCDP loses no information for Gaussian noise, loses at order s^2 for the Gaussian mixture, and only at order s^4 for Laplace noise; Figure 3(a) confirms these powers, with fitted slopes 2.00 and 4.00. For the discrete Gaussian, the loss is instead periodic in the order: zCDP is exact at orders where the shifted lattice realigns with $\mathbb{Z}$ and loses information in between (Figure 9, Appendix D.3). By (3.5), a workload of d queries has d times the one-query gap; Section 5.1 shows what this costs in noise variance.

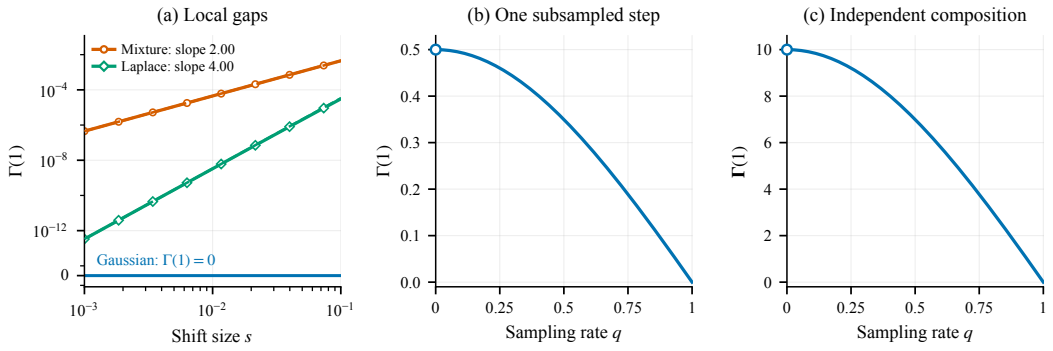


Figure 3: Information gaps at $t = 1$. (a) Gaussian noise has zero gap; the Gaussian-mixture ($m = 1$) and Laplace gaps are quadratic and quartic in s (solid: numerical; dashed: leading terms of Proposition 4.1; slopes fitted over $s \leq 0.01$). (b)–(c) Single-step and composed gaps of the subsampled Gaussian decrease in q ($n = 20$, $\rho_{\text{sum}} = 5$, $z = \sqrt{2}$; open circles are limits as $q \downarrow 0$). Appendix D.4.2 repeats these comparisons at $t = 1.5$.

4.2 Poisson subsampling

At every step, DP-SGD clips each per-example gradient to norm at most C, sums them over a Poisson subsample that includes each record independently with probability q, and adds Gaussian noise $N(0, z^2C^2I)$ with *noise multiplier* z (Abadi et al., 2016). Subsampling amplifies privacy. Finite-order RDP captures this amplification, whereas standard zCDP, whose guarantee must hold uniformly over all Rényi orders, does not retain it for the sampled Gaussian mechanism (Zhu & Wang, 2019; Bun et al., 2018). Under add-or-remove neighboring datasets, one step reduces to the sampled Gaussian mechanism (Mironov et al., 2019): $Q = N(0, z^2C^2I)$ and $P_q = (1-q)Q + qN(\mathbf{v}, z^2C^2I)$, where $\mathbf{v}$ is the clipped gradient of the extra record, with $\|\mathbf{v}\| = C$ in the worst case.

Proposition 4.2 (Poisson subsampling at a fixed zCDP budget)**.** *Fix $n \geq 1$, $C > 0$, and $\rho_{\text{sum}} > 0$, set $z_n^2 := n/(2\rho_{\text{sum}})$, and let (P_q, Q) be the one-step pair above with $z = z_n$ and $\|\mathbf{v}\| = C$.*

(a) *For $0 < q \leq 1$, we have*

$$K(t) = \log \mathbb{E}_{G\sim N(0,1)}\left[\left\{1 - q + q\exp\left(\frac{G}{z_n} - \frac{1}{2z_n^2}\right)\right\}^{t+1}\right],$$

$$\rho(P_q, Q) = \frac{\rho_{\text{sum}}}{n}, \qquad \text{and} \qquad \Gamma(t) = \frac{\rho_{\text{sum}}}{n}t(t+1) - K(t). \tag{4.2}$$

For every fixed $t > 0$, $\Gamma(t)$ is strictly decreasing in q on $(0, 1]$, equals 0 at $q = 1$, and tends to $\frac{\rho_{\text{sum}}}{n}t(t+1)$ as q tends to 0. At $t = 1$, we have $\Gamma(1) = \frac{2\rho_{\text{sum}}}{n} - \log\left\{1 + \left(e^{2\rho_{\text{sum}}/n} - 1\right)q^2\right\}$.

(b) *For n independent copies of this pair, the product pair has optimal zCDP parameter ρ_{sum}, and by* (3.5) *its gap is*

$$\mathbf{\Gamma}(t) = \rho_{\text{sum}}t(t+1) - \boldsymbol{K}(t) = n\Gamma(t),$$

which has the same monotonicity in q, vanishes at $q = 1$, and tends to $\rho_{\text{sum}}t(t+1)$ as q tends to 0.

At $q = 1$ there is no subsampling, each step is Gaussian, and zCDP loses no information. As q tends to 0, the exact RDP curve tends to zero while the zCDP parameter stays at ρ_{sum}, so the gap approaches its largest possible value $\rho_{\text{sum}}t(t+1)$: zCDP charges for privacy loss that does not occur (Figure 3(b)–(c)). This is one source of the utility differences in Section 5.2.

5 Experiments

We test whether the predictions of Section 4 persist when both notions are calibrated to the same privacy level $(\varepsilon_0, \delta_0) = (3, 10^{-5})$, and whether they affect utility, on an ACS table release and

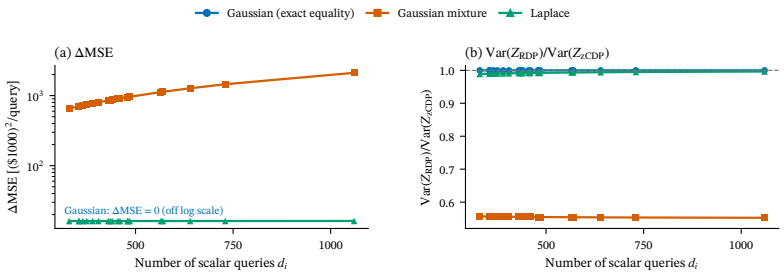


Figure 4: Noise variance required at $(\varepsilon_0, \delta_0) = (3, 10^{-5})$ on ACS-sized workloads. Left: $\Delta\mathrm{MSE}$ against the number of queries d_i (log scale; Gaussian is exactly 0 and omitted). Right: $\mathrm{Var}(Z_{\mathrm{RDP}})/\mathrm{Var}(Z_{\mathrm{zCDP}})$; smaller means that retaining the RDP curve saves more noise. Release-level distributions are in Figure 12.

on DP-SGD. Both accountants convert an RDP curve $U = (U_\alpha)_{\alpha>1}$ to (ε, δ)-DP with Balle et al. (2020b), optimized over the order:

$$\mathcal{C}_\delta[U] := \inf_{\alpha>1} \left\{ U_\alpha - \frac{\log\delta + \log\alpha}{\alpha - 1} + \log\frac{\alpha-1}{\alpha} \right\}. \tag{5.1}$$

The RDP accountant applies $\mathcal{C}_\delta$ to the exact curve, and the zCDP accountant to the zCDP envelope. Both apply the same conversion rule, isolating the effect of compression on noise calibration.

5.1 ACS-SIZED WORKLOAD

For each of the 20 states with the most counties, the workload consists of five county-level median earnings by educational attainment from the ACS (U.S. Census Bureau, 2020), giving $d_i = 5m_i$ queries for a state with m_i complete counties ($330 \le d_i \le 1060$), so by (3.5) the total gap is d_i times the one-query gap. This is a workload illustration built on published ACS aggregates, not a Census person-level DP mechanism (Appendix E.2). For each state and noise family (Gaussian, Gaussian-mixture, Laplace), we find the smallest per-query noise variance that certifies $(\varepsilon_0, \delta_0)$ under (5.1), $\mathrm{Var}(Z_{\mathrm{RDP}})$ and $\mathrm{Var}(Z_{\mathrm{zCDP}})$. The noise is centered and independent across queries, so the per-query MSE equals this variance; we report $\Delta\mathrm{MSE} := \mathrm{Var}(Z_{\mathrm{zCDP}}) - \mathrm{Var}(Z_{\mathrm{RDP}})$.

Result. For Gaussian noise the two calibrations coincide ($\Delta\mathrm{MSE} = 0$), since the Gaussian RDP curve equals its envelope. For the Gaussian mixture, retaining the RDP curve needs only 0.552–0.557 of the zCDP variance, about 45% lower noise variance, with $\Delta\mathrm{MSE}$ from 652.3 to 2115.7 in units of $(\$1000)^2$ per query. For Laplace noise the ratio is 0.989–0.997 and $\Delta\mathrm{MSE}$ is 16.170–16.205. This ordering matches the zero, quadratic, and quartic gaps of Proposition 4.1, although the ratios also depend on the expansion coefficients and the calibration. Monte Carlo releases with $20{,}000$ paired draws per state and family agree with the analytic $\Delta\mathrm{MSE}$ within 0.158%. Results at $(\varepsilon, \delta) = (5, 10^{-5})$ are in Appendix E.2.2.

5.2 DP-SGD

DP-SGD is where zCDP's simple bookkeeping is most attractive and, by Proposition 4.2, where it loses the most. We train a two-convolutional-layer CNN on Fashion-MNIST (Xiao et al., 2017) in Opacus (Yousefpour et al., 2021), with Poisson sampling, per-example clipping at $C = 1$, and the sampled-Gaussian RDP bound (Abadi et al., 2016; Mironov et al., 2019). For each sampling rate q and number of steps n, we compute the smallest noise multiplier that certifies $(\varepsilon_0, \delta_0)$ under each notion and train with it; a third, "clipping-only" run adds no noise. One sweep fixes $n = 500$ and varies $q \in \{0.005, 0.0075, 0.01, 0.02, 0.05\}$; the other fixes $q = 0.01$ and varies $n \in \{250, 500, 1000, 1500, 2000\}$. The privacy guarantees rest on per-step RDP bounds with adaptive composition; Proposition 4.2 only explains the trend. Each setting uses five seeds; within a seed, all runs share initialization, minibatches, and noise draws, so we report Student-t intervals over paired differences. Optimizer settings are in Appendix E.1.

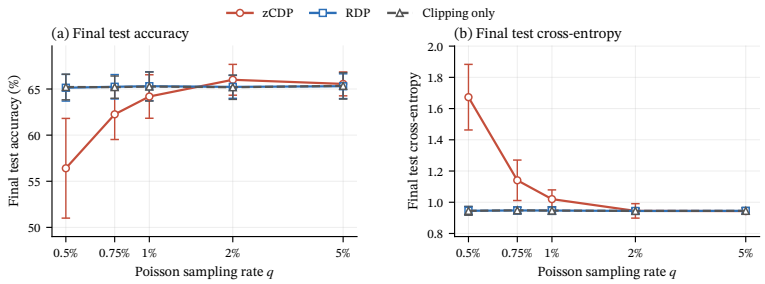


Figure 5: The sampling-rate sweep at $n = 500$ and $(\varepsilon, \delta) = (3, 10^{-5})$: final test accuracy and cross-entropy as q varies, with the noise recalibrated for each q. Means over five seeds; bars are 95% confidence intervals.

Sampling rate. With RDP calibration, test accuracy stays near 65.2% and cross-entropy near 0.95 across the q sweep (Figure 5), close to the clipping-only run. With zCDP calibration, accuracy rises from 56.408% at $q = 0.005$ to about 66% for $q \geq 0.02$ and cross-entropy falls from 1.673 to 0.943. At $q = 0.005$, RDP calibration reaches 65.138%, a paired gain of 8.73 ± 6.10 percentage points. The zCDP penalty shrinks as q grows, as Proposition 4.2 predicts. The seedwise log-q trend is significant for test loss ($p = 7.06 \times 10^{-4}$) and accuracy ($p = 0.0108$). Since changing q at fixed n also changes the batch size and the number of epochs (nq from 2.5 to 25), this trend is consistent with, but not a direct consequence of, Proposition 4.2.

Composition count. In Figure 2, RDP accuracy rises from 61.866% at $n = 250$ to 70.522% at $n = 2000$, within 0.08 percentage points of the clipping-only run in all nine settings. zCDP accuracy peaks at 66.354% at $n = 1000$ and falls to 64.600% at $n = 2000$, while its cross-entropy falls from 1.120 to 1.020 at $n = 500$ and then rises to 2.031. The paired test-loss difference grows with n (seedwise log-n trend $p = 2.9 \times 10^{-5}$). The paired accuracy difference is not monotone, and neither its trend ($p = 0.161$) nor the endpoint contrast ($p = 0.146$) is significant: each horizon is recalibrated, and optimization turns noise into accuracy nonlinearly, so cross-entropy is the cleaner measure here (training histories in Figure 16). Full results and sweeps at $(\varepsilon, \delta) = (5, 10^{-5})$ are in Appendix E.3.

6 Discussion and Limitations

Our results give a simple rule for choosing a privacy notion. The two one-sided (ε, δ)-DP curves, the trade-off functions, the PLD, and, when it is finite at one order, the exact RDP curve carry the same information about a pair (Theorem 2.1); choosing among them affects interpretation and computation, not information. The usual two-sided (ε, δ) curve loses only the direction, whereas zCDP is lossy even for a fixed pair. Its loss, measured by Γ, is zero for Gaussian noise and grows for mixture-type noise and under Poisson subsampling at small q (Figure 1). zCDP is therefore lossless for Gaussian noise without subsampling; otherwise, retaining one of the equivalent curves can reduce the noise needed for the same certified privacy.

Limitations. Theorem 2.1 concerns one fixed pair: a mechanism-level RDP bound takes the worst pair at each order and need not be the curve of any single pair, so the theorem does not apply to it directly. The gap isolates the RDP-to-zCDP step; the subsequent conversion to (ε, δ)-DP can be loose even when $\Gamma = 0$. The composition results assume independent components and add their zCDP parameters, and the local expansions hold for small shifts and bounded orders. The experiments cover three noise families, ACS-sized workloads, and one Fashion-MNIST model with five seeds, so they do not establish utility rankings for other datasets, optimizers, or deployments. Natural extensions include adaptive composition, mechanism-level analysis, and rigorous numerical evaluation of ρ when the maximizing order has no closed form.

AI USE STATEMENT

Generative AI tools were used to assist in checking mathematical arguments, reviewing analysis code, preparing figures, checking references, and editing the text for clarity. All theorems, proofs, experiments, and conclusions were developed and independently verified by the authors, who take full responsibility for the content of this paper.

ETHICS STATEMENT

The experiments use publicly available ACS aggregates and Fashion-MNIST; they do not collect new human-participant data. The ACS experiment is an illustration of noise calibration on published aggregate workloads, not a person-level privacy guarantee for Census respondents. The trained models are research artifacts for measuring utility, not models intended for deployment. The accountant comparisons hold for the stated output distributions, neighboring-dataset definitions, and conversion rules; deployment requires checking that these assumptions describe the actual mechanism.

REPRODUCIBILITY STATEMENT

The theorem assumptions appear in the main text and Appendices B–D; complete proofs appear in Appendices F–H. Appendix D.4.1 records numerical rules, grids, parameters, and resolution checks. ACS data preparation and paired noise experiments are specified in Appendix E.2; DP-SGD settings, calibration, seed pairing, and saved-result metadata are described in Appendices E.1 and E.3. No public repository for this study's code and saved experimental records is available at the time of writing.

REFERENCES

Martín Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pp. 308–318, 2016.

John M Abowd, Robert Ashmead, Ryan Cumings-Menon, Simson Garfinkel, Micah Heineck, Christine Heiss, Robert Johns, Daniel Kifer, Philip Leclerc, Ashwin Machanavajjhala, Brett Moran, William Sexton, Matthew Spence, and Pavel Zhuravlev. The 2020 Census Disclosure Avoidance System TopDown Algorithm. *Harvard Data Science Review*, 2022. doi: 10.1162/99608f92.529e3cb9. Special Issue 2.

Shahab Asoodeh, Jiachun Liao, Flavio P. Calmon, Oliver Kosut, and Lalitha Sankar. Three variants of differential privacy: Lossless conversion and applications. *IEEE Journal on Selected Areas in Information Theory*, 2(1):208–222, 2021. doi: 10.1109/JSAIT.2021.3054692.

Borja Balle, Gilles Barthe, and Marco Gaboardi. Privacy profiles and amplification by subsampling. *Journal of Privacy and Confidentiality*, 10(1), 2020a. doi: 10.29012/jpc.726.

Borja Balle, Gilles Barthe, Marco Gaboardi, Justin Hsu, and Tetsuya Sato. Hypothesis testing interpretations and Rényi differential privacy. In Silvia Chiappa and Roberto Calandra (eds.), *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 2496–2506. PMLR, 2020b.

Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography: 14th International Conference, TCC 2016-B, Part I*, volume 9985 of *Lecture Notes in Computer Science*, pp. 635–658. Springer, 2016. doi: 10.1007/978-3-662-53641-4_24.

Mark Bun, Cynthia Dwork, Guy N Rothblum, and Thomas Steinke. Composable and versatile privacy via truncated CDP. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pp. 74–86. Association for Computing Machinery, 2018. doi: 10.1145/3188745.3188946.

Clément L. Canonne, Gautam Kamath, and Thomas Steinke. The discrete Gaussian for differential privacy. In *Advances in Neural Information Processing Systems*, volume 33, pp. 15676–15688. Curran Associates, Inc., 2020.

Rishav Chourasia, Uzair Javaid, and Biplab Sikdar. Laplace transform interpretation of differential privacy. In *6th Symposium on Foundations of Responsible Computing (FORC 2025)*, volume 329 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 11:1–11:23. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025. doi: 10.4230/LIPIcs.FORC.2025.11. Extended version: arXiv:2411.09142.

Jinshuo Dong, Aaron Roth, and Weijie J. Su. Gaussian differential privacy. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84(1):3–37, 2022. doi: 10.1111/rssb.12454.

Vadym Doroshenko, Badih Ghazi, Pritish Kamath, Ravi Kumar, and Pasin Manurangsi. Connect the dots: Tighter discrete approximations of privacy loss distributions. *Proceedings on Privacy Enhancing Technologies*, 2022(4):552–570, 2022. doi: 10.56553/popets-2022-0122.

Cynthia Dwork. Differential privacy. In *Automata, Languages and Programming, 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II*, volume 4052 of *Lecture Notes in Computer Science*, pp. 1–12. Springer, 2006. doi: 10.1007/11787006_1.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4):211–407, 2014. doi: 10.1561/0400000042.

Badih Ghazi, Pritish Kamath, Ravi Kumar, and Pasin Manurangsi. Faster privacy accounting via evolving discretization. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 7470–7483. PMLR, 2022.

Juan Felipe Gomez, Bogdan Kulynych, Georgios Kaissis, Flavio P Calmon, Jamie Hayes, Borja Balle, and Antti Honkela. Gaussian dp for reporting differential privacy guarantees in machine learning. *arXiv preprint arXiv:2503.10945*, 2025.

Sivakanth Gopi, Yin Tat Lee, and Lukas Wutschitz. Numerical composition of differential privacy. In *Advances in Neural Information Processing Systems*, volume 34, pp. 11631–11642. Curran Associates, Inc., 2021.

Charlie Harrison and Pasin Manurangsi. Exact zCDP characterizations for fundamental differentially private mechanisms. In *7th Symposium on Foundations of Responsible Computing (FORC 2026)*, volume 368 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 3:1–3:18. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2026. doi: 10.4230/LIPIcs.FORC.2026.3.

Peter Kairouz, Sewoong Oh, and Pramod Viswanath. The composition theorem for differential privacy. *IEEE Trans. Inf. Theory*, 63(6):4037–4049, 2017.

Georgios Kaissis, Stefan Kolek, Borja Balle, Jamie Hayes, and Daniel Rueckert. Beyond the calibration point: Mechanism comparison in differential privacy. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*, volume 235 of *Proceedings of Machine Learning Research*, pp. 22840–22860. PMLR / OpenReview.net, 2024.

Daniel Kifer, John M. Abowd, Robert Ashmead, Ryan Cumings-Menon, Philip Leclerc, Ashwin Machanavajjhala, William Sexton, and Pavel Zhuravlev. Bayesian and frequentist semantics for common variations of differential privacy: Applications to the 2020 census. *arXiv preprint arXiv:2209.03310*, 2022.

Antti Koskela, Joonas Jälkö, and Antti Honkela. Computing tight differential privacy guarantees using FFT. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 2560–2569. PMLR, 2020.

Antti Koskela, Joonas Jälkö, Lukas Prediger, and Antti Honkela. Tight differential privacy for discrete-valued mechanisms and for the subsampled Gaussian mechanism using FFT. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 3358–3366. PMLR, 2021.

Ilya Mironov. Rényi differential privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pp. 263–275. IEEE, 2017. doi: 10.1109/CSF.2017.11.

Ilya Mironov, Kunal Talwar, and Li Zhang. Rényi differential privacy of the sampled Gaussian mechanism. *arXiv preprint arXiv:1908.10530*, 2019. doi: 10.48550/arXiv.1908.10530.

Anneliese Riess, Juan Felipe Gomez, Flavio du Pin Calmon, Julia Anne Schnabel, and Georgios Kaissis. Optimal conversion from Rényi differential privacy to f-differential privacy. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2602.04562. Accepted at ICML 2026; arXiv:2602.04562.

Buxin Su, Weijie J. Su, and Chendi Wang. The 2020 US Decennial Census is more private than you (might) think. *Proceedings of the National Academy of Sciences*, 122(45):e2500337122, 2025. doi: 10.1073/pnas.2500337122.

Buxin Su, Weijie Su, and Chendi Wang. A sieve-accelerated quadrature method for exact privacy accounting in the 2020 US Decennial Census. *arXiv preprint arXiv:2606.29835*, 2026.

Weijie J Su. A Statistical Viewpoint on Differential Privacy: Hypothesis Testing, Representation, and Blackwell's theorem. *Annual Review of Statistics and Its Application*, 12(1):157–175, 2025.

U.S. Census Bureau. S1501: Educational attainment (american community survey 5-year, 2020). https://data.census.gov/table/ACSST5Y2020.S1501, 2020. Table ID: S1501; Version: 2016–2020 5-year estimates.

Yu-Xiang Wang, Borja Balle, and Shiva Prasad Kasiviswanathan. Subsampled Rényi differential privacy and analytical moments accountant. In Kamalika Chaudhuri and Masashi Sugiyama (eds.), *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 1226–1235. PMLR, 2019.

Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017. doi: 10.48550/arXiv.1708.07747.

Ashkan Yousefpour, Igor Shilov, Alexandre Sablayrolles, Davide Testuggine, Karthik Prasad, Mani Malek, John Nguyen, Sayan Ghosh, Akash Bharadwaj, Jessica Zhao, Graham Cormode, and Ilya Mironov. Opacus: User-friendly differential privacy library in PyTorch. *arXiv preprint arXiv:2109.12298*, 2021. doi: 10.48550/arXiv.2109.12298.

Yuqing Zhu and Yu-Xiang Wang. Poission subsampled Rényi differential privacy. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 7634–7642. PMLR, 2019.

Yuqing Zhu, Jinshuo Dong, and Yu-Xiang Wang. Optimal accounting of differential privacy via characteristic function. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 4782–4817. PMLR, 2022.

A Related work

A statistical view of differential privacy places hypothesis testing and Blackwell informativeness at the center of privacy comparison (Su, 2025). Within this view, privacy profiles and hypothesis-testing trade-off functions are connected through convex duality (Dong et al., 2022; Balle et al., 2020b). Gopi et al. (2021) formalize privacy loss random variables as an equivalent reparameterization of a full privacy curve. Zhu et al. (2022) introduce the characteristic function of the privacy-loss random variable as a lossless notion and derive conversions to and from privacy profiles, trade-off functions, and PLDs. Wang et al. (2019) observe that a finite privacy-loss cumulant-generating function identifies the PLD, while Chourasia et al. (2025) derive a Laplace-transform correspondence between an extended privacy profile and the Rényi-divergence curve under absolute-continuity conditions. A complementary line of work studies optimal black-box conversions from reported RDP bounds to (ε, δ)-DP or f-DP guarantees (Asoodeh et al., 2021; Riess et al., 2026).

Relation to prior equivalence results. Theorem 2.1 unifies these connections into an identifiability result for the standard real-order exact RDP curve, with an explicit regularity condition and counterexamples clarifying the roles of directional information and finite exponential moments. For arbitrary probability measures, including singular pairs, finiteness of $D_{\alpha_0}(P\|Q)$ at one real order $\alpha_0 > 1$ makes the full RDP curve determine the extended PLD, the hypothesis-testing trade-off curve, and both one-sided (ε, δ)-DP curves. Recovery uses only real-order RDP values, through exponential tilting and moment-generating-function uniqueness, without mutual absolute continuity or complex-order reports. Neither qualification can simply be dropped: the pointwise maximum of the one-sided curves can lose f-DP and RDP information, while an everywhere-infinite RDP curve can correspond to different one-sided (ε, δ)-DP curves. To our knowledge, prior work has not developed an order-resolved divergence measure of this RDP-to-zCDP compression, nor characterized its local cumulant structure and accumulation under composition.

B Additional results for Section 2

This section collects the standard definitions and counterexamples for Section 2. Proofs are given in Appendix F.

B.1 Standard privacy definitions

Let P and Q be the output distributions on neighboring datasets.

f-DP and the trade-off function $T_{P,Q}$. For testing P against Q, $T_{P,Q}(a)$ is the smallest type II error among tests whose type I error is at most a: $T_{P,Q}(a) := \inf_\phi\{1 - \mathbb{E}_Q[\phi] : \mathbb{E}_P[\phi] \le a\}$ for $0 \le a \le 1$, where ϕ ranges over randomized tests with values in $[0, 1]$. A mechanism is f-DP if $T_{P,Q} \ge f$ for every ordered neighboring pair (Dong et al., 2022).

Privacy loss and RDP. Writing p and q for the densities or probability mass functions of P and Q, the privacy loss is $L = \log\{p(X)/q(X)\}$ for $X \sim P$. Its distribution is the privacy-loss distribution (PLD) (Zhu et al., 2022). If an output is possible under P but impossible under Q, its privacy loss is $+\infty$. The definition for general output distributions is given in Appendix F. For order $\alpha = 1 + t > 1$, the Rényi divergence and the cumulant-generating function (CGF) of L satisfy $K(t) := \log \mathbb{E}_P[e^{tL}]$ and $D_{1+t}(P\|Q) = K(t)/t$. These quantities may be infinite; we set $K(0) = 0$. An $(\alpha, \varepsilon_\alpha)$-RDP guarantee bounds $D_\alpha(P\|Q)$ by ε_α for every ordered neighboring pair (Mironov, 2017).

zCDP. A mechanism is $\widetilde{\rho}$-zCDP if $D_\alpha(P\|Q) \le \widetilde{\rho}\alpha$ for every $\alpha > 1$ and every ordered neighboring pair (Bun & Steinke, 2016). For a fixed pair, the smallest such parameter is $\rho(P, Q) := \sup_{\alpha>1} D_\alpha(P\|Q)/\alpha$. Thus zCDP replaces the RDP curve by a single number.

B.2 Why both one-sided (ε, δ)-DP curves are needed

Without an exchange symmetry, the two one-sided curves may differ. Their pointwise maximum can lose which ordering generated each supporting line of the f-DP curve. The next full-support example

shows this even when all likelihood ratios are bounded and every RDP value is finite. For $\gamma \geq 1$, write $H_\gamma(P\|Q) = \sup_A\{P(A) - \gamma Q(A)\}$, so that $H_\gamma(P\|Q) = \delta_{P\|Q}(\log\gamma)$.

Example B.1 (The pointwise maximum of the two one-sided curves can lose ordering information)**.** Let $c = 6/11$. Define two full-support probability mass functions on $\mathbb{Z}$ by

$$q_A(k) = q_B(k) = c4^{-k} \quad (k \geq 0),$$

and, for $j \geq 1$, by

$$q_A(-j) = c2^{-(j+1)}, \qquad q_B(-j) = c3^{-j}.$$

For $i \in \{A, B\}$, let $Q_i = \mathrm{Law}(Z_i)$ and $P_i = \mathrm{Law}(Z_i + 1)$ when Z_i has mass function q_i. For $\gamma = e^\varepsilon \geq 1$, the two forward hockey-stick divergences agree:

$$H_\gamma(P_i\|Q_i) = \frac{4c}{3}\left(1 - \frac{\gamma}{4}\right)_+.$$

The reverse divergences are

$$H_\gamma(Q_A\|P_A) = c\left(1 - \frac{\gamma}{4}\right)_+ + \frac{c}{2}\left(1 - \frac{\gamma}{2}\right)_+,$$
$$H_\gamma(Q_B\|P_B) = \frac{3c}{2}\left(1 - \frac{\gamma}{3}\right)_+.$$

The common forward one-sided (ε, δ)-DP curve dominates both reverse one-sided (ε, δ)-DP curves for every $\gamma \geq 1$. Hence the two pairs have the same two-sided (ε, δ)-DP curve. Nevertheless, at $\gamma = 3$ their reverse divergences are $3/22$ and 0, respectively, so their ordered f-DP curves T_{P_A,Q_A} and T_{P_B,Q_B} differ. Their RDP curves also differ at order two:

$$\exp(D_2(P_A\|Q_A)) = \frac{265}{88}, \qquad \exp(D_2(P_B\|Q_B)) = 3.$$

All forward and reverse likelihood ratios are bounded by 4. The strongest symmetric f-DP guarantee obtained from the two-sided (ε, δ)-DP curve is the same for the two pairs; taking the pointwise maximum discards which one-sided (ε, δ)-DP curve generates each part of the ordered f-DP curve.

Example B.2 (Finite mean privacy loss is insufficient)**.** Write $\zeta(w) = \sum_{k=1}^\infty k^{-w}$ and $\mathrm{Li}_w(z) = \sum_{k=1}^\infty z^k/k^w$. On $\Omega = \{\partial, 1, 2, \ldots\}$, for $w > 2$, define

$$P_w\{k\} = \frac{k^{-w}}{\zeta(w)}, \qquad P_w\{\partial\} = 0,$$
$$Q_w\{k\} = P_w\{k\}e^{-k}, \qquad Q_w\{\partial\} = 1 - \frac{\mathrm{Li}_w(e^{-1})}{\zeta(w)}.$$

Then Q_w is a probability measure, $P_w \ll Q_w$, and

$$\mathbb{E}_{P_w} L_w = \frac{\zeta(w-1)}{\zeta(w)}, \qquad \delta_{P_w\|Q_w}(1) = \frac{1}{\zeta(w)}\sum_{k=2}^\infty \frac{1 - e^{1-k}}{k^w}.$$

For $w = 3$ and $w = 4$, respectively,

$$\mathbb{E}_{P_3} L_3 = 1.36843\ldots, \qquad \delta_{P_3\|Q_3}(1) = 0.124865\ldots,$$
$$\mathbb{E}_{P_4} L_4 = 1.11063\ldots, \qquad \delta_{P_4\|Q_4}(1) = 0.053061\ldots.$$

Both mean privacy losses are finite, and the two (ε, δ)-DP curves are different. Nevertheless, $D_\alpha(P_w\|Q_w) = +\infty$ for every $\alpha > 1$ and either value of w. Thus finite mean privacy loss is insufficient for recovery from the RDP curve.

C Additional results for Section 3

C.1 From the gap to privacy bounds

Fix a mutually absolutely continuous pair (P, Q) with CGF K and finite optimal zCDP parameter ρ. For $c_t = t^t/(1+t)^{1+t}$, the fixed-order RDP-to-(ε, δ) conversion gives

$$\delta_{P\|Q}(\varepsilon) \leq c_t \exp\{K(t) - t\varepsilon\}, \qquad t > 0. \tag{C.1}$$

This is the conversion of Balle et al. (2020b); its derivation is in Appendix G.2.

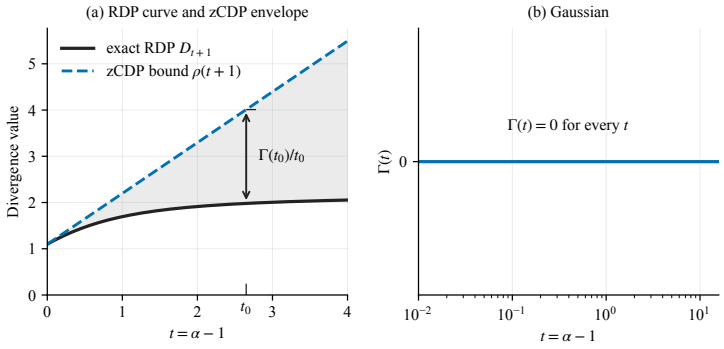


Figure 6: The information gap. Left: $\Gamma(t)/t$ is the vertical distance between the exact RDP curve and its zCDP envelope; here $P = (9, 3, 3, 1)/16$ and $Q = (1, 3, 3, 9)/16$ on four points, $\rho = \log 3$, and $t_0 = 2.65$. Right: for Gaussian noise ($\rho = 1/2$), $\Gamma(t) = 0$ for every $t > 0$.

The ε bound at a fixed δ. Fix $\delta \in (0, 1)$ and keep the output pair, including its noise scale, unchanged. Define

$$F_{\mathrm{R}}(t) := \frac{K(t) + \log c_t - \log \delta}{t}, \tag{C.2}$$

$$F_{\mathrm{Z}}(t) := \rho(t+1) + \frac{\log c_t - \log \delta}{t} = F_{\mathrm{R}}(t) + \frac{\Gamma(t)}{t}. \tag{C.3}$$

Minimizing these objectives over $t > 0$ gives the ε upper bounds obtained by applying the same RDP-to-(ε, δ) conversion to the exact curve and its zCDP envelope. Proposition 3.2 gives

$$\frac{\Gamma(t_{\mathrm{Z}})}{t_{\mathrm{Z}}} \le \varepsilon_{\mathrm{Z}} - \varepsilon_{\mathrm{R}} \le \frac{\Gamma(t_{\mathrm{R}})}{t_{\mathrm{R}}}. \tag{C.4}$$

The relevant quantity for calibration is therefore $\Gamma(t)/t$ near the optimizing orders, rather than the gap at an arbitrary reference order. In particular, a large gap at t_{Z} certifies a nonnegligible compression cost, whereas a uniform upper bound around t_{R} certifies that the cost is small. Obtaining such an interval and bound without computing the full RDP curve requires additional mechanism-specific analysis. The proposition compares certificates at a fixed noise scale; translating it into a noise-variance bound additionally requires control of how the certificate changes with that scale.

Using the local expansion. Let $G_s(t)$ be the three terms preceding $\mathcal{R}_s(t)$ in (3.3). Under the assumptions of Theorem 3.1 and Proposition 3.2, if $0 < t_{\mathrm{R}} \le T$, then

$$0 \le \varepsilon_{\mathrm{Z}} - \varepsilon_{\mathrm{R}} \le \frac{G_s(t_{\mathrm{R}})}{t_{\mathrm{R}}} + \frac{M_T}{120}(t_{\mathrm{R}} + t_{\mathrm{R}}^4)|s|^5. \tag{C.5}$$

This follows by substituting (3.4) into (C.4). The condition $t_{\mathrm{R}} \le T$ specifies the orders for which this remainder estimate applies.

The δ bound at a fixed ε. For any valid zCDP upper bound $\widetilde{\rho} \ge \rho(P, Q)$, define

$$\delta_{\mathrm{zCDP}}(\varepsilon; \widetilde{\rho}) := \inf_{t>0} c_t \exp\{\widetilde{\rho} t(t+1) - t\varepsilon\}. \tag{C.6}$$

For $\widetilde{\rho} > 0$, the minimizer $t_\varepsilon(\widetilde{\rho})$ is unique and satisfies

$$\log \frac{t_\varepsilon(\widetilde{\rho})}{1 + t_\varepsilon(\widetilde{\rho})} + \widetilde{\rho}\{2t_\varepsilon(\widetilde{\rho}) + 1\} - \varepsilon = 0. \tag{C.7}$$

When $\rho > 0$, taking $\widetilde{\rho} = \rho$ gives

$$\delta_{P\|Q}(\varepsilon) \le e^{-\Gamma(t_\varepsilon)} \delta_{\mathrm{zCDP}}(\varepsilon; \rho), \qquad t_\varepsilon = t_\varepsilon(\rho). \tag{C.8}$$

Thus, at the zCDP optimizing order, retaining the exact RDP curve reduces the δ bound by a factor $e^{-\Gamma(t_\varepsilon)}$. Appendix G.2 proves these statements.

C.2 Further consequences of the local expansion

For the mechanism calculations, it is useful to divide out powers of the shift. For $s \neq 0$, define

$$U_s := \frac{L_s - \mathbb{E}_{P_s} L_s}{s}, \qquad A_s := \frac{2\{\rho(s) - \mathbb{E}_{P_s} L_s\}}{s^2} \geq 0. \tag{C.9}$$

The nonnegativity follows from $\lim_{t\downarrow 0} K(t)/\{t(t+1)\} = K'(0) = \mathbb{E}_{P_s} L_s$ and $\rho(s) = \sup_{t>0} K(t)/\{t(t+1)\}$. These definitions give $\rho(s) - \mathbb{E}_{P_s} L_s = s^2 A_s/2$, $\kappa_3(L_s) = s^3 \kappa_3(U_s)$, and $\kappa_4(L_s) = s^4 \kappa_4(U_s)$.

Recall $\boldsymbol{K}$, ρ_{sum}, and $\boldsymbol{\Gamma}$ from Section 3.2.

Theorem 3.3 states that, for every $t > 0$,

$$tD_{t+1}(P^{(n)} \| Q^{(n)}) = \boldsymbol{K}(t), \qquad \boldsymbol{\Gamma}(t) = \rho_{\mathrm{sum}} t(t+1) - tD_{t+1}(P^{(n)} \| Q^{(n)}). \tag{C.10}$$

A component with $P_i = Q_i$ has $K_i(t) = \rho_i = \Gamma_i(t) = 0$.

For independent components, let $(P_i, Q_i) = (P_{i,s_i}, Q_{i,s_i})$ and write $\mathbf{s} = (s_1, \ldots, s_n)$. Suppose each component family satisfies Theorem 3.1 at its nonzero shift s_i, with fifth-derivative constant $M_{i,T}$. Define A_{i,s_i} and U_{i,s_i} by (C.9) for that pair, and let $\mathcal{I} = \{i : s_i \neq 0\}$. At $s_i = 0$, the two component distributions agree and $K_i(t) = \rho_i = 0$. Theorem 3.3 gives

$$\begin{aligned} \boldsymbol{\Gamma}(t) = \frac{t(t+1)}{2} \sum_{i \in \mathcal{I}} A_{i,s_i} s_i^2 - \frac{t^2(t+1)}{6} \sum_{i \in \mathcal{I}} \kappa_3(U_{i,s_i}) s_i^3 \\ + \frac{t^2(1-t^2)}{24} \sum_{i \in \mathcal{I}} \kappa_4(U_{i,s_i}) s_i^4 + \mathcal{R}_{\mathbf{s}}(t), \end{aligned} \tag{C.11}$$

$$|\mathcal{R}_{\mathbf{s}}(t)| \leq \frac{t^2 + t^5}{120} \sum_{i \in \mathcal{I}} M_{i,T} |s_i|^5, \qquad 0 < t \leq T. \tag{C.12}$$

The optimal zCDP parameter of the product pair is

$$\rho_\star := \sup_{t>0} \frac{\boldsymbol{K}(t)}{t(t+1)} \leq \rho_{\mathrm{sum}}.$$

Consequently,

$$\boldsymbol{\Gamma}(t) = (\rho_{\mathrm{sum}} - \rho_\star) t(t+1) + \{\rho_\star t(t+1) - \boldsymbol{K}(t)\}. \tag{C.13}$$

Both terms are nonnegative. The first records the difference between adding the component parameters and optimizing the parameter of the product pair; it vanishes for identical components. The second is the gap between the product pair's RDP curve and its optimal zCDP envelope.

D Additional results for Section 4

D.1 Allocation under a fixed zCDP composition parameter

We consider n independent applications of the same one-dimensional additive-noise mechanism. Let $Z_1, \ldots, Z_n$ be independent copies of a real-valued random variable Z, and, for shift magnitudes $r_i \geq 0$, set

$$Q_i = \mathrm{Law}(Z_i), \qquad P_{i,r_i} = \mathrm{Law}(Z_i + r_i), \qquad i = 1, \ldots, n. \tag{D.1}$$

We call this setting *homogeneous*: the variables Z_i are different independent noise draws with the same distribution, although their shift magnitudes may differ.

Recall the single-component quantities $K(t)$, $\rho(r)$, and $\Gamma(t)$ from Section 2 and (3.1). Fix $\rho_{\mathrm{sum}} > 0$, and assume that $\rho(0) = 0$ and that $r \mapsto \rho(r)$ is continuous and strictly increasing on a range whose image contains $[0, \rho_{\mathrm{sum}}]$. Its inverse on this interval is denoted by

$$r = \psi(x), \qquad \rho\big(\psi(x)\big) = x, \qquad 0 \leq x \leq \rho_{\mathrm{sum}}. \tag{D.2}$$

At $r = 0$, we use $K(t) = \Gamma(t) = 0$. Set $\rho_i = \rho(r_i)$ and $r_i = \psi(\rho_i)$. We study allocations satisfying

$$\rho_i \geq 0, \qquad \sum_{i=1}^{n} \rho_i = \rho_{\mathrm{sum}}. \tag{D.3}$$

Each ρ_i is the optimal zCDP parameter of its component pair. For fixed t, define the calibrated single-pair gap function

$$\Phi_t(x) := \Gamma(t)|_{r=\psi(x)}, \qquad 0 \leq x \leq \rho_{\mathrm{sum}}. \tag{D.4}$$

Equivalently, $\Phi_t(x) = xt(t+1) - K(t)$, where $K(t)$ is evaluated for the single pair with $r = \psi(x)$. Substituting $r_i = \psi(\rho_i)$ into (3.5) gives

$$\boldsymbol{\Gamma}(t) = \rho_{\mathrm{sum}} t(t+1) - \boldsymbol{K}(t) = \sum_{i=1}^{n} \Phi_t(\rho_i). \tag{D.5}$$

Because every component has the same noise distribution, the same functions $\rho(r)$, $\psi(x)$, and $\Phi_t(x)$ apply to all i. Jensen's inequality applied to (D.5) gives the following allocation rule.

Proposition D.1 (Allocation of a fixed total zCDP parameter)**.** *Fix $t > 0$ and consider all allocations satisfying* (D.3)*.*

(a) *If Φ_t is convex on $[0, \rho_{\mathrm{sum}}]$, then every feasible allocation satisfies*

$$n\Phi_t\Big(\frac{\rho_{\mathrm{sum}}}{n}\Big) \leq \boldsymbol{\Gamma}(t) \leq \Phi_t(\rho_{\mathrm{sum}}). \tag{D.6}$$

Hence the uniform zCDP allocation

$$\rho_i = \frac{\rho_{\mathrm{sum}}}{n}, \qquad r_i = \psi\left(\frac{\rho_{\mathrm{sum}}}{n}\right), \qquad i = 1, \ldots, n,$$

minimizes the gap. Every permutation of the concentrated allocation

$$(\rho_1, \ldots, \rho_n) = (\rho_{\mathrm{sum}}, 0, \ldots, 0), \qquad (r_1, \ldots, r_n) = (\psi(\rho_{\mathrm{sum}}), 0, \ldots, 0),$$

maximizes it.

(b) *If Φ_t is concave on $[0, \rho_{\mathrm{sum}}]$, then*

$$\Phi_t(\rho_{\mathrm{sum}}) \leq \boldsymbol{\Gamma}(t) \leq n\Phi_t\Big(\frac{\rho_{\mathrm{sum}}}{n}\Big),$$

so concentration minimizes the gap and uniform allocation maximizes it.

If the relevant function is strictly convex or strictly concave, the uniform extremum is unique, and the concentrated extrema are exactly the n permutations displayed above.

Proposition D.1 is pointwise in t: it ranks allocations by the tightness of the standard composed zCDP bound at one RDP order, and the same allocation need not be optimal at every order.

Figure 7 illustrates both alternatives in Proposition D.1. In both panels, $t = 1$, $n = 8$, $\rho_{\mathrm{sum}} = 1/2$, and

$$\begin{aligned} &\rho_1(\omega) = \frac{\rho_{\mathrm{sum}}}{n}\{1 + (n-1)\omega\}, \quad \rho_i(\omega) = \frac{\rho_{\mathrm{sum}}}{n}(1-\omega), \quad i = 2, \ldots, n, \\ &r_i(\omega) = \psi\big(\rho_i(\omega)\big), \qquad\qquad 0 \leq \omega \leq 1. \end{aligned} \tag{D.7}$$

Thus $\sum_i \rho_i(\omega) = \rho_{\mathrm{sum}}$ for every ω, $\omega = 0$ is the uniform zCDP allocation, and $\omega = 1$ is the concentrated shift allocation $\{\psi(\rho_{\mathrm{sum}}), 0, \ldots, 0\}$.

For unit-scale Laplace noise, $\rho(r) = r + e^{-r} - 1$, and the proof verifies that Φ_1 is strictly convex. The composed gap therefore increases from 0.037202015 at the uniform allocation to 0.193535993 at the concentrated allocation. For the symmetric Gaussian mixture with $m = 1$ in Proposition 4.1, $\rho(r) = r^2/2$, so $\psi(x) = \sqrt{2x}$. Centered finite differences of the exact CGF are negative on the displayed grid at two Gauss–Hermite resolutions, consistently with concavity of Φ_1 on $[0, 1/2]$. Along the displayed allocation path, the composed gap decreases from 0.435179700 at the uniform allocation to 0.342190807 at a concentrated allocation.

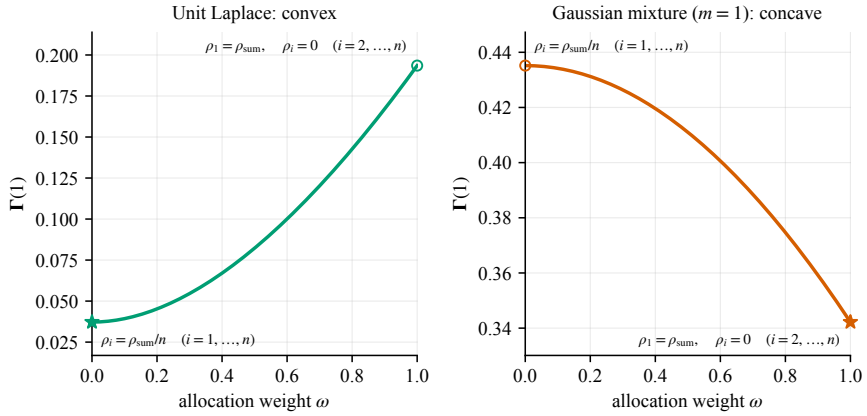


Figure 7: Exact standard-composition gaps $\mathbf{\Gamma}(1)$ are plotted as a fixed total zCDP parameter is reallocated across components. Concentration has opposite effects for the two mechanisms: it increases the Laplace gap analytically but decreases the Gaussian-mixture gap along the displayed numerically checked path. Both panels impose $\sum_i \rho(r_i) = \rho_{\text{sum}} = 1/2$ and use the allocation path (D.7), where $\omega = 0$ is uniform and $\omega = 1$ is concentrated. Left: strict convexity of the unit-scale Laplace function Φ_1 makes the uniform allocation the unique minimum and concentration the maximum. Right: for symmetric Gaussian-mixture noise with $m = 1$, centered finite differences are negative on the numerical grid, and the uniform and concentrated endpoints give the largest and smallest computed gaps. Numerical evaluation details are given in Appendix D.4.1. The $t = 1.5$ counterpart appears in Appendix D.4.2.

D.2 Dependence on tail parameters for single location pairs

We next isolate the effect of a tail parameter from the effect of composition by studying one translated ordered pair at a time. For Student-t_ν noise, the optimal zCDP parameter agrees locally with the mean privacy loss and the first nonzero gap term is quartic. For generalized Gaussian noise, it also contains the nonnegative difference $\rho_\beta(s) - \mathbb{E}_{P_s} L_s$. The Student coefficient is nonmonotone in ν, so tail weight alone does not determine the gap. The Student-t expansion below follows by combining endpoint localization with the local expansion theorem; its coefficient is then evaluated explicitly.

Proposition D.2 (Student-t_ν location pair)**.** *Let*

$$f_\nu(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}, \qquad \nu \geq 1,$$

and let Z have density f_ν, $P_s = \mathrm{Law}(Z + s)$, and $Q = \mathrm{Law}(Z)$. Write $\rho_\nu(s) := \rho(P_s, Q)$. Then $\rho_\nu(s) < \infty$ for every $s \in \mathbb{R}$. For every fixed $\nu \geq 1$, one has $\rho_\nu(s) = \mathbb{E}_{P_s} L_s$ for all sufficiently small nonzero s. Moreover, for every fixed $T > 0$, uniformly for $0 < t \leq T$ and all sufficiently small nonzero s,

$$\Gamma(t) = c_{4,\nu}[t(t+1)]^2 s^4 + O(|s|^6), \qquad c_{4,\nu} := \frac{3(\nu+1)^2(\nu^2+4\nu-1)}{4\nu(\nu+3)^2(\nu+5)(\nu+7)}. \tag{D.8}$$

The coefficient has a unique maximum at $\nu = 6.013052641\ldots$ and satisfies $c_{4,\nu} = 3/(4\nu) + O(\nu^{-2}) \to 0$ as $\nu \to \infty$.

The generalized-Gaussian expansion below follows from its score-cumulant calculation and the local expansion theorem; the Laplace and Gaussian endpoints are checked against their exact transforms.

Proposition D.3 (Generalized Gaussian location pair)**.** *Let*

$$f_\beta(x) \propto \exp(-|x|^\beta/\beta), \qquad 1 \leq \beta \leq 2,$$

and let Z have density f_β, $P_s = \mathrm{Law}(Z + s)$, and $Q = \mathrm{Law}(Z)$. Write $\rho_\beta(s) := \rho(P_s, Q)$ and define

$$\chi_\beta := \frac{\beta^{4-4/\beta}}{24\Gamma(1/\beta)^2}\left\{3\Gamma(2-1/\beta)^2 - \Gamma(4-3/\beta)\Gamma(1/\beta)\right\} \geq 0.$$

The optimal pairwise zCDP parameter $\rho_\beta(s)$ is finite for every $s \in \mathbb{R}$. For every fixed $\beta \in [1, 2]$ and $T > 0$, uniformly for $0 < t \le T$ and all sufficiently small nonzero s,

$$\begin{aligned}\Gamma(t) = t(t+1)\{\rho_\beta(s) - \mathbb{E}_{P_s} L_s\} \\ + \chi_\beta[t(t+1)]^2 s^4 + o(|s|^4).\end{aligned} \tag{D.9}$$

At $\beta = 1$,

$$\rho_1(s) = \mathbb{E}_{P_s} L_s = |s| + e^{-|s|} - 1, \qquad \chi_1 = \frac{1}{12},$$

recovering Laplace noise. At $\beta = 2$,

$$\rho_2(s) = \mathbb{E}_{P_s} L_s = \frac{s^2}{2}, \qquad \chi_2 = 0,$$

and $\Gamma(t) = 0$ identically for this Gaussian endpoint.

For $1 < \beta < 2$, the finite-shift difference $\rho_\beta(s) - \mathbb{E}_{P_s} L_s$ is not available in closed form. Since $\rho_\beta(s) \ge \mathbb{E}_{P_s} L_s$, for every fixed $T > 0$, uniformly for $0 < t \le T$ and all sufficiently small nonzero s,

$$\Gamma(t) \ge \chi_\beta[t(t+1)]^2 s^4 + o(|s|^4). \tag{D.10}$$

Figure 8 compares the complete single-pair gaps at $t = 1$ and $s = 0.2$. For the generalized Gaussian family, we evaluate $\rho_\beta(s)$ by numerically maximizing $K(u)/[u(u+1)]$ over $u > 0$ for each displayed pair, including the $u \downarrow 0$ endpoint. The dashed curves show the quartic local-expansion approximations supplied by Propositions D.2 and D.3. The full generalized Gaussian expansion in (D.9) also contains $2\{\rho_\beta(s) - \mathbb{E}_{P_s} L_s\}$ at $t = 1$; the numerical optimization finds this difference to be zero to numerical precision at every displayed value of β.

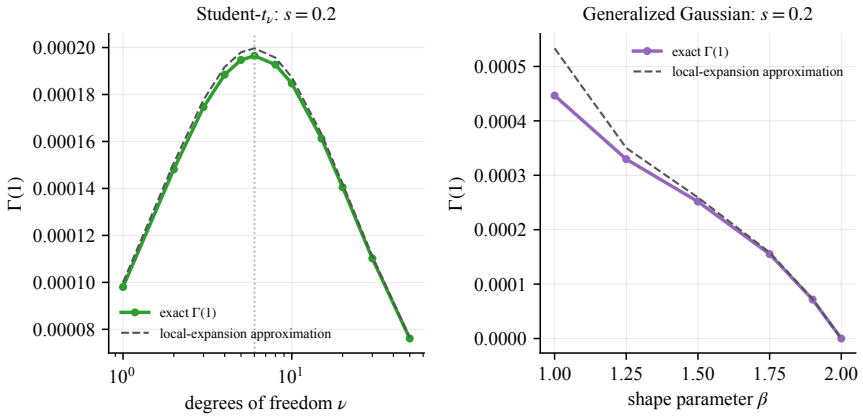


Figure 8: Exact single-pair gaps $\Gamma(1)$ are compared with quartic local-expansion approximations as two tail parameters vary, fixing $t = 1$ and $s = 0.2$. The Student-t gap is nonmonotone in ν, whereas the generalized Gaussian gap decreases to zero at the Gaussian endpoint $\beta = 2$, so tail parameters do not induce a universal monotone ordering of the gap. Left: for Student-t_ν pairs, the dashed curve is $4c_{4,\nu}s^4$ from (D.8), and the vertical dotted line marks the maximizer of $c_{4,\nu}$. Right: for generalized Gaussian pairs, $\rho_\beta(s)$ is obtained by maximizing the normalized CGF for each β, and the dashed curve is the local-expansion approximation $4\chi_\beta s^4$ from (D.10). In both formulas, $4 = [t(t+1)]^2$ at $t = 1$; the dashed curves are small-s approximations, not finite-s bounds. Appendix D.4.2 gives the same comparison at $t = 1.5$.

D.3 Discrete and lattice mechanisms

We finally consider lattice-valued noise. Discrete Laplace privacy loss has finitely many values and hence a finite-sum CGF. The discrete Gaussian gap is periodic in t, vanishes when the shifted lattices align, and is exponentially small as $\sigma \to \infty$. The formulas below give both gaps exactly.

Proposition D.4 (Exact gaps for discrete Laplace and discrete Gaussian noise)**.** *Let θ be integer-valued and fix neighboring datasets D, D' for which $s := \theta(D) - \theta(D') \neq 0$ and $r := |s| \in \{1, 2, \ldots\}$. Because both noise laws below are centrally symmetric, Corollary F.1 lets us use the positive representative of each ordered location pair. As in the continuous families above, let L_r denote the privacy loss of each displayed pair. The following formulas hold for every $t > 0$ and extend continuously to $t = 0$, where the gap is zero.*

(a) Fix $a > 0$, let Z have discrete Laplace mass proportional to $e^{-a|z|}$, put $Q = \mathrm{Law}(Z)$, $P_r = \mathrm{Law}(Z + r)$, and $\lambda = e^{-a}$. Then

$$\begin{aligned} K(t) &= \log\left[\frac{1-\lambda}{1+\lambda}\left\{\frac{e^{tar} + \lambda^r e^{-tar}}{1-\lambda} + \sum_{j=1}^{r-1} \lambda^j e^{ta(r-2j)}\right\}\right], \\ \rho(r) &= \mathbb{E}_{P_r} L_r = a\left\{r + (\lambda^r - 1)\,\mathrm{csch}(a)\right\}, \\ \Gamma(t) &= a\left\{r + (\lambda^r - 1)\,\mathrm{csch}(a)\right\} t(t+1) - K(t). \end{aligned} \tag{D.11}$$

The finite sum is empty when $r = 1$.

(b) For discrete Gaussian noise with scale $\sigma > 0$, let Z have mass function

$$\mathbb{P}(Z = k) = \frac{e^{-k^2/(2\sigma^2)}}{\Theta_\sigma(0)}, \qquad \Theta_\sigma(u) := \sum_{k \in \mathbb{Z}} e^{-(k-u)^2/(2\sigma^2)}.$$

Put $Q = \mathrm{Law}(Z)$ and $P_r = \mathrm{Law}(Z + r)$. Then

$$\begin{aligned} K(t) &= \frac{r^2}{2\sigma^2} t(t+1) + \log\frac{\Theta_\sigma((1+t)r)}{\Theta_\sigma(0)}, \\ \rho(r) &= \mathbb{E}_{P_r} L_r = \frac{r^2}{2\sigma^2}, \\ \Gamma(t) &= \log\frac{\Theta_\sigma(0)}{\Theta_\sigma((1+t)r)} = 4e^{-2\pi^2\sigma^2} \sin^2(\pi(1+t)r) + O(e^{-4\pi^2\sigma^2}), \end{aligned} \tag{D.12}$$

uniformly in t as $\sigma \to \infty$. The gap is $1/r$-periodic in t and vanishes exactly when $(1+t)r \in \mathbb{Z}$.

Figure 9 plots the discrete Gaussian and discrete Laplace gaps $\Gamma(t)$ over the same range of the order offset $t = \alpha - 1$. The discrete Laplace curves are computed from the exact transform in Proposition D.4. Fixing the discrete Gaussian scale at $\sigma = 0.60$ keeps the periodic correction and its exact zeros visible.

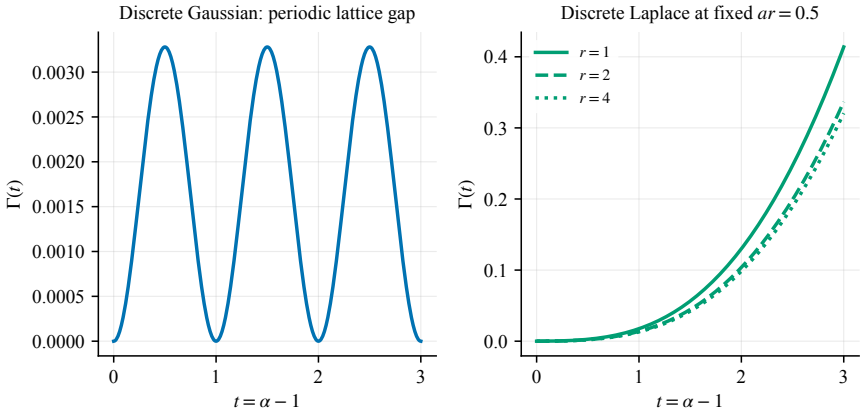


Figure 9: The exact zCDP–RDP gaps $\Gamma(t)$ of two lattice mechanisms are compared over the same range of the order offset. The periodic exact zeros for discrete Gaussian noise and the positive discrete Laplace gaps for $t > 0$ show how the accuracy of zCDP can vary with the RDP order. Left: discrete Gaussian noise uses $r = 1$ and $\sigma = 0.60$, with exact zeros whenever $(1 + t)r \in \mathbb{Z}$. Right: discrete Laplace noise uses the labeled shifts $r \in \{1, 2, 4\}$ at fixed $ar = 0.5$, so $a = 0.5/r$; the curves are computed from Proposition D.4. Both panels use $0 \le t \le 3$ but retain separate vertical scales.

D.4 Additional numerical results

D.4.1 Computational setup

The analytical figures in Sections 3 and 4 and Appendix D are deterministic. Gaussian, Laplace, and lattice curves use the closed forms in Propositions 4.1 and D.4. Gaussian-mixture expectations use 192-node Gauss–Hermite quadrature; Student-t_ν expectations use 2048-node Gauss–Legendre quadrature after $x = \sqrt{\nu}\tan\theta$; and generalized-Gaussian transforms use adaptive quadrature split at $-s$ and 0. For the last family, we maximize $K(u)/[u(u+1)]$ over $0 < u \le 32$, including the $u \downarrow 0$ endpoint. The large-order truncation is checked separately from the asymptotic statement in Proposition D.3. Specifically, for $1 < \beta < 2$, $r = |s|$, and $0 < \zeta < 1$, the bound

$$\frac{K(u)}{u(u+1)} \le B_{\beta,r,\zeta}(u) := \frac{ur^\beta + (ur)^\beta(\beta-1)^{\beta-1}/(\beta\zeta^{\beta-1}) - \beta^{-1}\log(1-\zeta)}{u(u+1)}$$

follows from $L_s \le r|Z|^{\beta-1} + r^\beta$, Young's inequality, and $\mathbb{E}\exp\{\zeta|Z|^\beta/\beta\} = (1-\zeta)^{-1/\beta}$. For fixed ζ, every term of $B_{\beta,r,\zeta}(u)$ is nonincreasing when $u \ge (\beta-1)/(2-\beta)$. At the displayed $r = 0.2$ and $\beta \in \{1.25, 1.5, 1.75, 1.9\}$, this threshold is at most 9. Choosing ζ to minimize $B_{\beta,r,\zeta}(32)$ gives an upper bound below the numerically evaluated endpoint $K'(0) \le \rho(s)$ in every case; the smallest margin is 2.7296×10^{-4}. These parameter-specific tail checks justify the cutoff used in the plots. The endpoints $\beta = 1, 2$ use the exact Laplace and Gaussian formulas. Exponential moments use log-sum-exp arithmetic where needed.

The remaining figure-specific settings are as follows. Figure 6 uses the captioned four-point pair and, for its Gaussian panel, $\rho = 1/2$ on 100 log-spaced values $10^{-2} \le t \le 10^{1.2}$. Figure 3(a) fixes $t = 1$ and uses 61 log-spaced shifts $10^{-3} \le s \le 10^{-1}$, with $m = 1$ and unit-scale Laplace noise; slopes are fitted over $s \le 0.01$. Figure 3(b)–(c) uses 1001 values of $q \in [0, 1]$ with $(n, \rho_{\text{sum}}, \|\mathbf{v}\|/C, t) = (20, 5, 1, 1)$ and plots the positive-q right limit at $q = 0$. Figure 7 uses 241 values of $\omega \in [0, 1]$ with $(n, \rho_{\text{sum}}, t) = (8, 1/2, 1)$; repeating the Gaussian-mixture calculation with 384 rather than 192 Hermite nodes leaves every interior centered second difference negative, providing numerical evidence for concavity along this path. Figure 8 fixes $(t, s) = (1, 0.2)$ and uses $\nu \in \{1, 2, 3, 4, 5, 6.013052641, 8, 10, 15, 20, 30, 50\}$ and $\beta \in \{1, 1.25, 1.5, 1.75, 1.9, 2\}$. Finally, Figure 9 uses 361 values on $0 \le t \le 3$, with $(r, \sigma) = (1, 0.60)$ for discrete Gaussian noise and $r \in \{1, 2, 4\}$ at $ar = 0.5$ for discrete Laplace noise. Solid and dashed curves denote the complete gaps and the stated local-expansion terms, respectively; dotted lines mark the special parameter values identified in the captions.

D.4.2 RESULTS AT AN ALTERNATIVE REPORTING ORDER

The fixed-order displays in Section 4 use $t = 1$ (Rényi order $\alpha = 2$). Figures 10 and 11 repeat every such comparison at $t = 1.5$ ($\alpha = 2.5$), with all mechanism parameters and numerical grids unchanged. Figure 9 already spans $0 \le t \le 3$, so the lattice examples require no duplicate fixed-order panel.

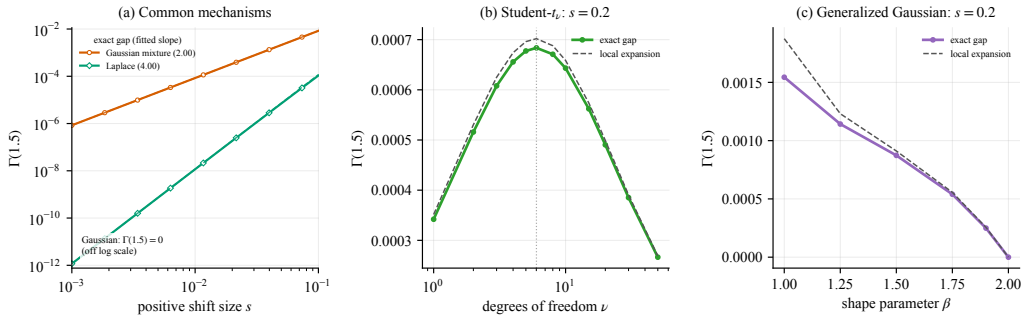


Figure 10: The single-pair mechanism and tail-parameter comparisons are recomputed at $t = 1.5$. The zero, quadratic, and quartic local regimes and the two tail-parameter patterns from the $t = 1$ figures remain visible. (a) Gaussian, $m = 1$ Gaussian-mixture, and unit-Laplace gaps over $10^{-3} \le s \le 10^{-1}$; solid curves are exact numerical gaps and dashed curves are the local leading terms, with fitted slopes shown in parentheses. (b) Student-t_ν and (c) generalized Gaussian gaps at $s = 0.2$, with dashed quartic local-expansion approximations.

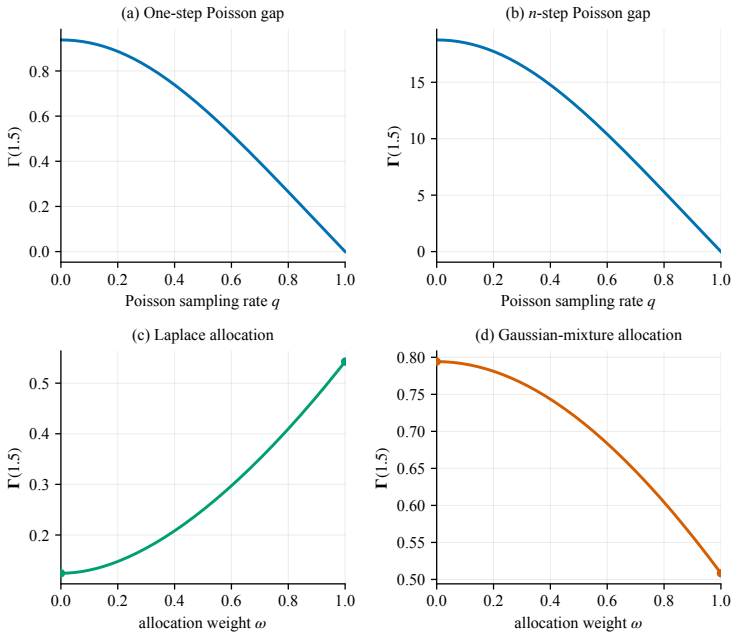


Figure 11: The Poisson-subsampling and fixed-zCDP-allocation comparisons are recomputed at $t = 1.5$. The Poisson gaps still decrease from their positive-q right limits to zero, while concentrating the allocation still increases the Laplace gap and decreases the computed Gaussian-mixture gap. (a) One-step $\Gamma(1.5)$ and (b) n-step $\mathbf{\Gamma}(1.5) = n\Gamma(1.5)$ for $n = 20$ and $\rho_{\mathrm{sum}} = 5$. (c) Unit-Laplace and (d) $m = 1$ Gaussian-mixture composition gaps along (D.7) for $n = 8$ and $\rho_{\mathrm{sum}} = 1/2$. The Poisson values at $q = 0$ are right-hand limits of the positive-q family.

The fitted common-mechanism slopes are 2.000 and 3.997 for the Gaussian mixture and Laplace families. The Gaussian endpoint remains exactly zero. The Student-t_ν curve remains nonmonotone, while the generalized Gaussian curve decreases to its exact zero at $\beta = 2$.

For the noninteger Poisson order, we evaluate the expectation in (4.2) by a 192-node Gauss–Hermite rule and repeat it with 384 nodes; the largest CGF discrepancy is 7.11×10^{-15}. The exact composition agrees with $n\Gamma(1.5)$ to 1.78×10^{-15}. The allocation calculations reuse the exact Laplace transform and the two Gaussian-mixture quadrature resolutions. In the displayed ρ grid, the minimum Laplace curvature is 1.15 and the maximum Gaussian-mixture curvature is -1.41, preserving the two allocation directions.

E Additional results for Section 5

E.1 Numerical methods and reproducibility

The analytical figures are deterministic; every quadrature rule, grid, and parameter value, together with two-resolution agreement checks (largest CGF discrepancy 7.11×10^{-15}), is recorded in Appendix D.4.1. The ACS workload uses published ACS 5-year aggregates (U.S. Census Bureau, 2020); state selection, completeness filtering ($1{,}961$ of $2{,}098$ rows in the top-20 states), the scalar shift $\Delta = \$1{,}000$, base-noise definitions, and the $20{,}000$ shared paired draws per state and family are in Appendix E.2. The DP-SGD experiments use the standard Fashion-MNIST $60{,}000/10{,}000$ split (Xiao et al., 2017) with fixed normalization and no augmentation; a two-convolution CNN with a 64-unit hidden layer and no batch normalization; flat clipping $C = 1$; SGD with initial learning rate 0.05, no momentum, weight decay 10^{-4}, and a common 2000-step cosine schedule; five seeds per cell; the grids $q \in \{0.005, 0.0075, 0.01, 0.02, 0.05\}$ at $n = 500$ and $n \in \{250, 500, 1000, 1500, 2000\}$ at $q = 0.01$; and targets $(\varepsilon, \delta) = (3, 10^{-5})$ and, in the appendix, $(5, 10^{-5})$. Within each cell and seed all methods share the initialization and Poisson stream and the private pair shares the standard-normal stream, so comparisons are exactly paired. Maximum calibration errors in ε were 8.67×10^{-10} and 1.52×10^{-9} at $\varepsilon = 3$ and $\varepsilon = 5$, respectively. Per-cell endpoint metadata appear in Tables 3 and 4. All data used are already public (ACS published aggregates and Fashion-MNIST).

The following subsections give the metadata and alternative-privacy experiments associated with Sections 5.1 and 5.2.

E.2 Additional results for the ACS noise-addition experiment

E.2.1 Experimental setup and metadata

The source file contains $3{,}221$ county and county-equivalent rows. After excluding the District of Columbia and Puerto Rico, the 50 states contain $3{,}142$ rows, of which $2{,}902$ have numeric values in all five target columns. The top 20 states by raw county count contain $2{,}098$ rows; $1{,}961$ are jointly complete, so 137 (6.53%) are excluded without imputation. The source tokens and counts for each column are recorded in the accompanying CSV. Table 1 lists the selected states and the resulting dimensions. Ranking uses raw county count with alphabetical state name as the tie-break. In the paper's notation, m_i is the number of jointly complete counties in state i, and $d_i = 5m_i$ is the number of county-by-education queries in its workload.

Earnings are retained in dollars. The scalar ordered-pair shift is fixed at $\Delta = \$1{,}000$ for every component. For a candidate dollar scale λ, the standardized shift used in the one-component transforms is $s = \Delta/\lambda = 1000/\lambda$. The centered base noises are $N(0, 1)$; the equal-weight mixture $N(-1, 1)/2 + N(1, 1)/2$; and unit-scale Laplace. The corresponding variances are 1, 2, and 2. These conventions determine the one-query CGF and optimal zCDP parameter used in Section 5.1.

The underlying metadata record contains, for every selected state and family, both calibrated shifts and dollar scales, active orders, achieved privacy, absolute RDP and zCDP MSE, analytic $\Delta\mathrm{MSE}$, and Monte Carlo summaries. The earnings values themselves cancel algebraically from the squared-error difference and serve only to define and label the workload.

Table 1: The 20 ACS-defined state workloads used throughout the paper. Here m_i is the number of counties or county-equivalents with all five earnings entries numeric, and $d_i = 5m_i$ is the number of scalar queries in the composed workload.

Rank	State	m_i	d_i	Rank	State	m_i	d_i
1	TX	212	1060	11	NE	71	355
2	GA	146	730	12	IN	91	455
3	VA	128	640	13	OH	87	435
4	KY	113	565	14	MN	86	430
5	MO	114	570	15	MI	81	405
6	KS	88	440	16	MS	78	390
7	IL	96	480	17	OK	75	375
8	NC	96	480	18	AR	73	365
9	IA	97	485	19	WI	71	355
10	TN	92	460	20	AL	66	330

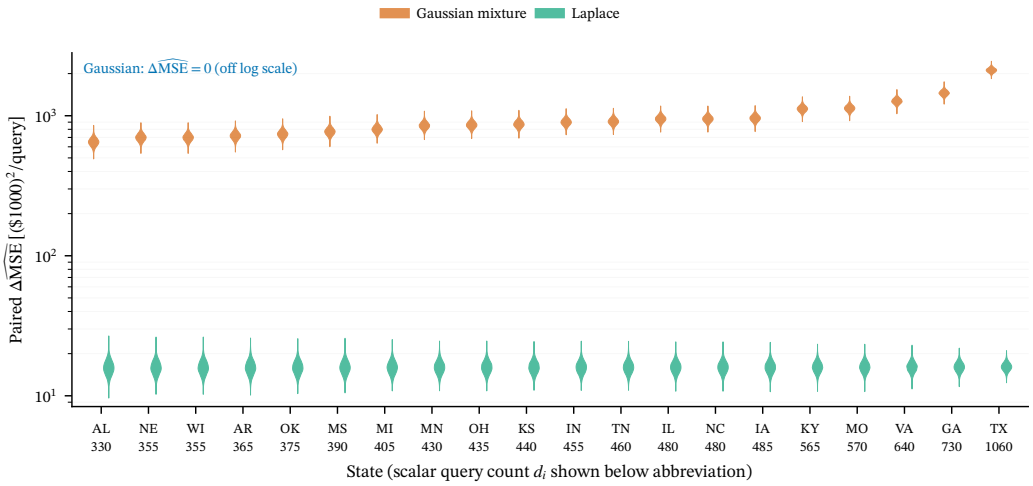


Figure 12: The paired $\widehat{\Delta\text{MSE}}$ distributions of the two non-Gaussian families are compared across the 20 main-text ACS workloads. Their separation reproduces the ordering in Figure 4: Gaussian-mixture values exceed Laplace values. The horizontal axis gives states ordered by d_i; the logarithmic vertical axis gives $\widehat{\Delta\text{MSE}}$ in units of $(\$1000)^2$ per query. At each state, each family violin contains 20,000 paired draws. Gaussian is identically zero and is omitted from the logarithmic axis. The violin widths show the distribution of single-release outcomes, not confidence intervals.

E.2.2 Results at an alternative privacy target

We repeat the ACS experiment at $(\varepsilon, \delta) = (5, 10^{-5})$, changing only the privacy target. The workloads, queries, mechanisms, and 20,000 paired draws are unchanged, so Figures 13 and 14 are directly parallel to Figures 4 and 12. The Gaussian equality and the ordering of the two non-Gaussian families persist, while the absolute MSE differences are smaller at the weaker privacy target. Table 2 records the ranges underlying the two displays.

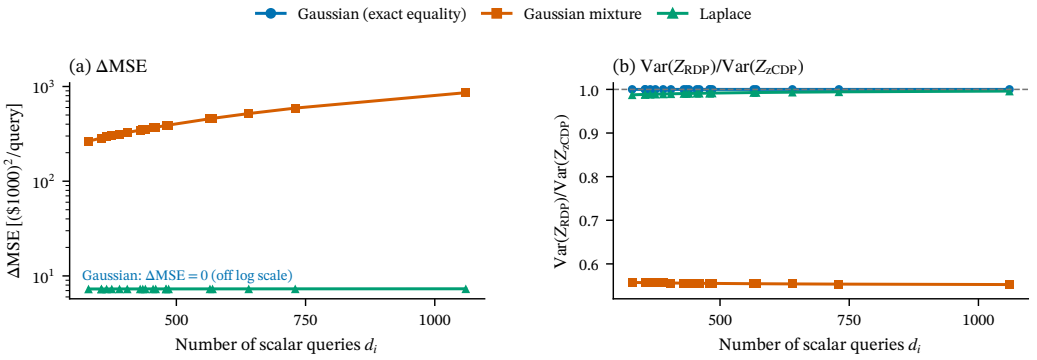


Figure 13: The ACS accounting-to-utility comparison is repeated at $(\varepsilon, \delta) = (5, 10^{-5})$ on the same 20 state workloads. Gaussian RDP-to-zCDP compression remains exact, and the other two families preserve the ordering at $(3, 10^{-5})$, consistently with the quadratic–quartic local hierarchy. Left: analytic $\Delta\mathrm{MSE}$ versus $d_i = 5m_i$; Gaussian is exactly zero and is omitted from the logarithmic axis. Right: $\mathrm{Var}(Z_{\mathrm{RDP}})/\mathrm{Var}(Z_{\mathrm{zCDP}})$ versus d_i, with Gaussian at one.

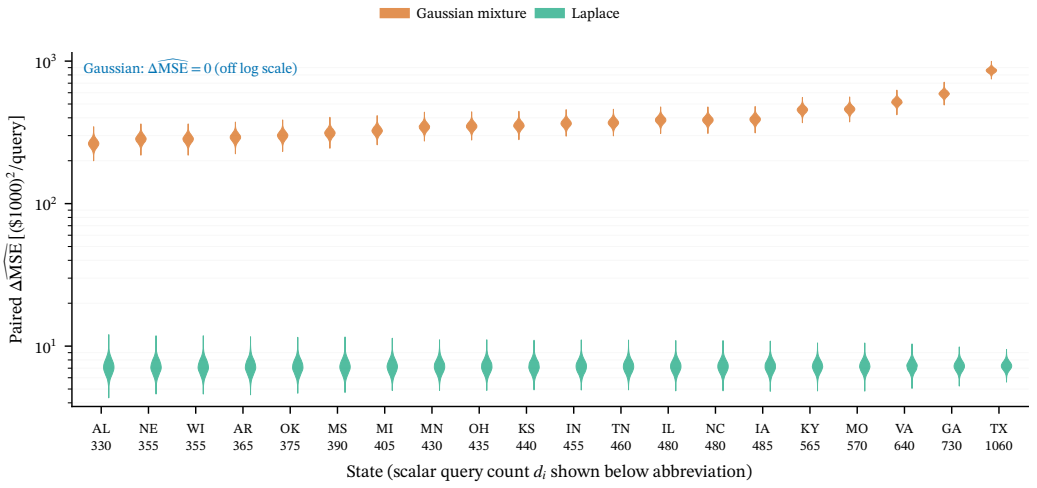
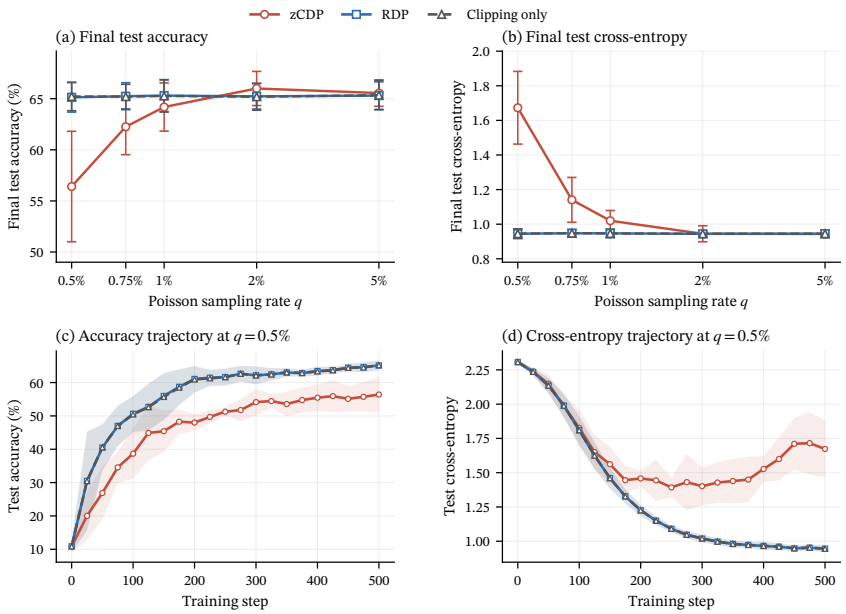


Figure 14: The paired ACS release-level MSE differences are shown at $(\varepsilon, \delta) = (5, 10^{-5})$ for the same top-20 state workloads. The family ordering is unchanged: Gaussian-mixture differences exceed Laplace differences. The horizontal axis orders states by d_i, the logarithmic vertical axis reports $\Delta\widehat{\mathrm{MSE}}$ in $(\$1000)^2$ per query, and each violin contains 20,000 paired draws. Gaussian is identically zero and is omitted.

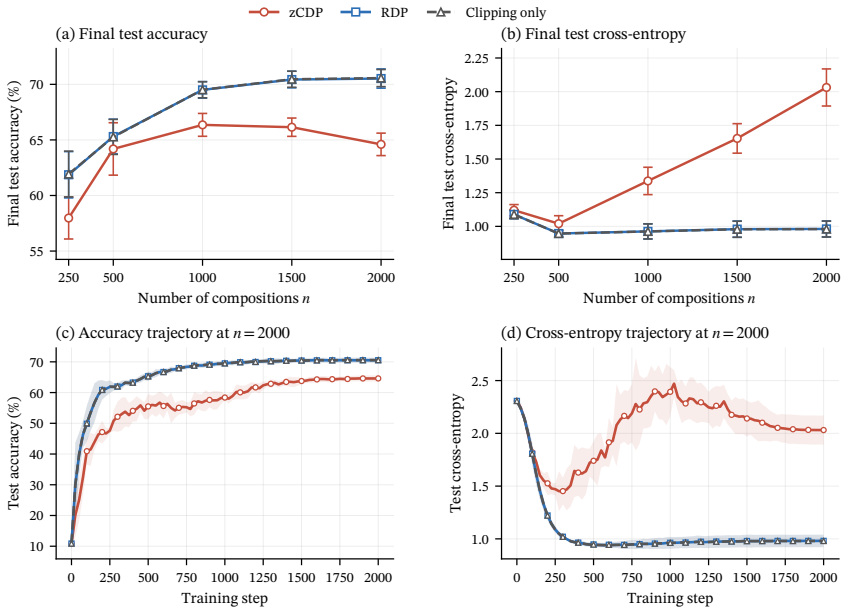
Table 2: ACS endpoint metadata at $(\varepsilon, \delta) = (5, 10^{-5})$. Ranges are over the 20 workloads in Table 1; $\Delta\mathrm{MSE}$ is analytic and is measured in $(\$1000)^2$ per query.

Noise family	$\Delta\mathrm{MSE}$ range	$\mathrm{Var}(Z_{\mathrm{RDP}})/\mathrm{Var}(Z_{\mathrm{zCDP}})$ range
Gaussian	0	1
Gaussian mixture	265.0–860.6	0.553–0.557
Laplace	7.279–7.313	0.988–0.996



Final points and trajectory curves are five-seed means; bars and bands are 95% Student-t confidence intervals. All panels include the clipping-only baseline.

Figure 15: The sampling-rate experiment at $(\varepsilon, \delta) = (3, 10^{-5})$, all cells at $n = 500$ training steps. (a)–(b): final test accuracy and test cross-entropy of the RDP-calibrated, zCDP-calibrated, and zero-noise clipping runs as q varies. (c)–(d): the corresponding three-method histories at the low-sampling endpoint $(q, n) = (0.005, 500)$. Curves are five-seed means; bars and bands are 95% confidence intervals.



Final points and trajectory curves are five-seed means; bars and bands are 95% Student-t confidence intervals. All panels include the clipping-only baseline.

Figure 16: The composition-count experiment at $(\varepsilon, \delta) = (3, 10^{-5})$ and $q = 0.01$. (a)–(b): final test accuracy and cross-entropy as n varies. (c)–(d): training histories for $n = 2000$, using the noise calibrated for $n = 2000$ throughout. Curves are five-seed means; bars and bands are 95% confidence intervals.

E.3 Additional results for the DP-SGD subsampling and composition experiment

E.3.1 Experimental setup and metadata

Fashion-MNIST images use the standard train/test split and fixed normalization, without augmentation. The network has two convolution and max-pooling stages followed by a 64-unit hidden layer and contains no batch normalization. The expected Poisson batch size is qN; Opacus clips each example to flat norm $C = 1$, adds Gaussian noise once per training step, and divides by that expected batch size. All horizons use the same 2000-step cosine learning-rate schedule. Test metrics and a fixed deterministic training probe are evaluated every 25 steps; minibatch metrics are recorded at every training step. The endpoint histories are included in Figures 15 and 16. Table 3 records the nine unique final cells underlying those figures and the paired RDP–zCDP summaries used in the text. The RDP and zero-noise clipping mean accuracies differ by at most 0.08 percentage points, while the zCDP response varies with the calibrated noise burden.

Table 3: Complete controlled DP-SGD endpoint metadata. The shared $(q, n) = (0.01, 500)$ cell is listed once, and nq is the expected number of epochs. In the upper block, accuracies are percentages and L_Z, L_R, and L_0 are final test cross-entropies for zCDP, RDP, and the zero-noise clipping control. The lower block reports paired RDP–zCDP means and 95% confidence intervals; ΔAcc is in percentage points.

q	n	nq	Acc_Z	Acc_R	Acc_0	L_Z	L_R	L_0
0.0050	500	2.5	56.408	65.138	65.218	1.673	0.945	0.945
0.0075	500	3.75	62.256	65.244	65.204	1.141	0.947	0.948
0.0100	500	5	64.188	65.304	65.270	1.020	0.947	0.946
0.0200	500	10	66.000	65.232	65.192	0.945	0.944	0.945
0.0500	500	25	65.554	65.304	65.342	0.943	0.945	0.945
0.0100	250	2.5	57.972	61.866	61.928	1.120	1.089	1.089
0.0100	1000	10	66.354	69.502	69.512	1.337	0.963	0.962
0.0100	1500	15	66.144	70.432	70.460	1.653	0.979	0.978
0.0100	2000	20	64.600	70.522	70.564	2.031	0.981	0.980

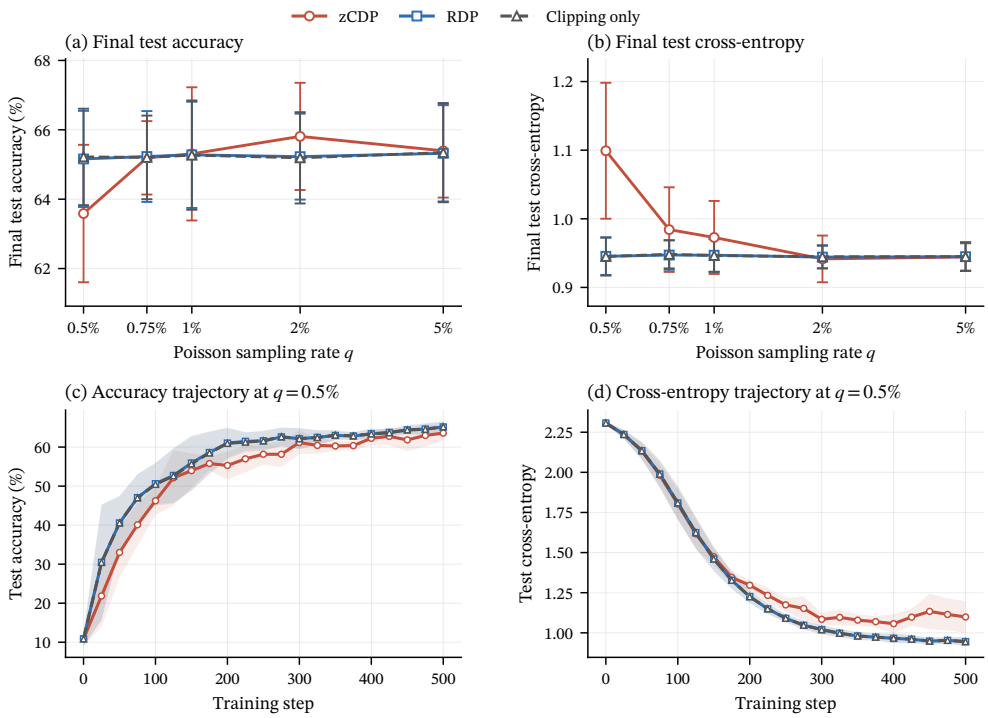
q	n	ΔL_{test}	ΔAcc
0.0050	500	0.728 ± 0.231	8.73 ± 6.10
0.0075	500	0.193 ± 0.123	2.99 ± 2.60
0.0100	500	0.073 ± 0.050	1.12 ± 2.92
0.0200	500	0.000 ± 0.039	-0.77 ± 1.44
0.0500	500	-0.001 ± 0.006	-0.25 ± 0.36
0.0100	250	0.031 ± 0.035	3.89 ± 1.77
0.0100	1000	0.375 ± 0.104	3.15 ± 1.15
0.0100	1500	0.673 ± 0.102	4.29 ± 1.31
0.0100	2000	1.050 ± 0.110	5.92 ± 1.48

All 115 scheduled trajectories completed and produced all 135 logical results. Initialization and Poisson-stream hashes agree within every triplet, and the private-pair standard-normal hashes agree. At $q = 0.01$, shorter horizons are exact prefixes of the longest clipping trajectory. The maximum calibration error is 8.67×10^{-10} in ε; refining the order grid changes a noise multiplier by at most 7.38×10^{-7} relatively, no optimum lies on a search boundary, and the $q = 1$ Gaussian anchor has relative error 9.46×10^{-10}. The runs used Opacus 1.6.0 and PyTorch 2.11.0 on an NVIDIA A100-SXM4-80GB. Fixed pseudorandom streams make the comparisons reproducible on the public dataset.

E.3.2 Results at an alternative privacy target

We repeat the controlled DP-SGD experiment at $(\varepsilon, \delta) = (5, 10^{-5})$, changing only the privacy target. The architecture, sweep grids, five seeds, and paired random streams remain unchanged. Figures 17

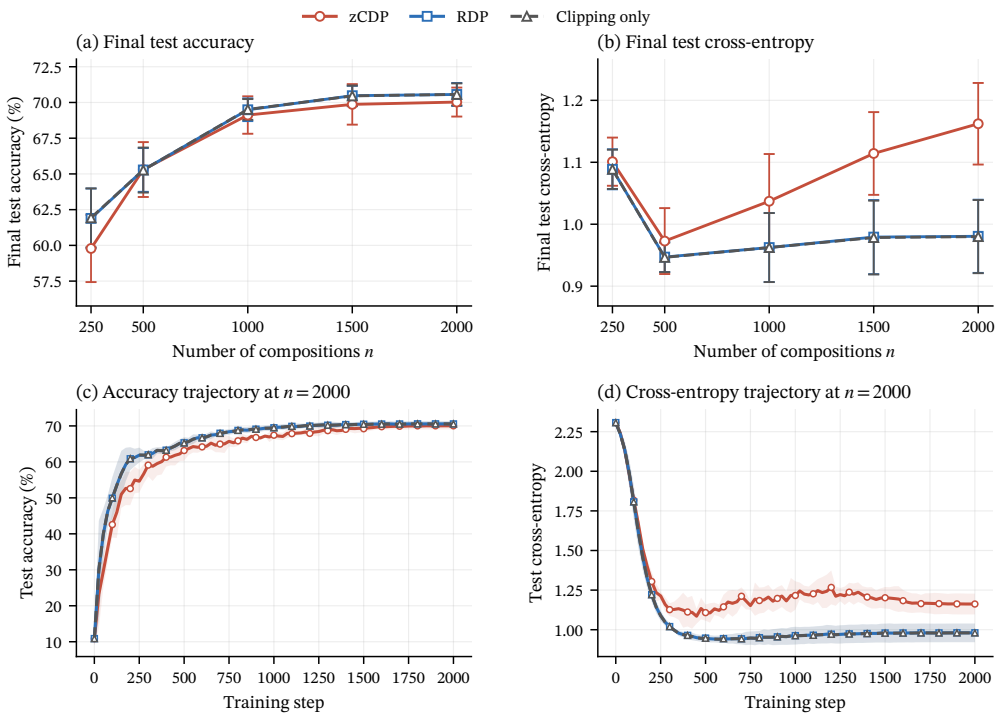
and 18 therefore repeat the complete four-panel layouts of Figures 15 and 16. At the weaker privacy target, the RDP–zCDP utility differences are smaller, but the test-cross-entropy separation again shrinks toward zero across q and increases with n.



Final points and trajectory curves are five-seed means; bars and bands are 95% Student-t confidence intervals. All panels include the clipping-only baseline.

Figure 17: The controlled sampling-rate sweep is repeated at $(\varepsilon, \delta) = (5, 10^{-5})$. The low-$q$ zCDP penalty is smaller than at the main-text target, while the zCDP curve again approaches the RDP and clipping-only curves as q increases. All cells use $n = 500$; panels (a)–(b) report final accuracy and test cross-entropy, and panels (c)–(d) report their trajectories at $(q, n) = (0.005, 500)$. All panels include the three methods; bars and bands are five-seed 95% confidence intervals.

Table 4 gives all endpoint values. The paired test-loss difference decreases from 0.154 at $q = 0.005$ to approximately zero at $q = 0.05$ and increases from 0.012 at $n = 250$ to 0.182 at $n = 2000$. The corresponding seedwise log-trends have $p = 0.0107$ and $p = 0.0152$. The accuracy trends are not significant at this target ($p = 0.106$ for q and $p = 0.361$ for n), so the alternative run supports the loss-based pattern without implying monotonicity of accuracy.



Final points and trajectory curves are five-seed means; bars and bands are 95% Student-t confidence intervals. All panels include the clipping-only baseline.

Figure 18: The controlled composition-count sweep is repeated at $(\varepsilon, \delta) = (5, 10^{-5})$. The paired test-cross-entropy separation again grows with n, whereas accuracy remains a noisier, nonmonotone downstream response. All cells use $q = 0.01$; panels (a)–(b) report final accuracy and test cross-entropy, and panels (c)–(d) report the longest-cell trajectories at $(q, n) = (0.01, 2000)$. All panels include the three methods. The trajectories use the noise calibrated for $n = 2000$ throughout, so their value at training step 500 is not the separately calibrated $n = 500$ endpoint in panels (a)–(b). Bars and bands are five-seed 95% confidence intervals.

Table 4: Complete controlled DP-SGD endpoint metadata at $(\varepsilon, \delta) = (5, 10^{-5})$. Notation and units match Table 3; the lower block reports paired means and 95% confidence intervals.

q	n	nq	Acc_Z	Acc_R	Acc_0	L_Z	L_R	L_0
0.0050	500	2.5	63.586	65.160	65.218	1.099	0.945	0.945
0.0075	500	3.75	65.192	65.230	65.204	0.984	0.947	0.948
0.0100	500	5	65.306	65.278	65.270	0.973	0.947	0.946
0.0200	500	10	65.810	65.226	65.192	0.942	0.944	0.945
0.0500	500	25	65.394	65.322	65.342	0.944	0.945	0.945
0.0100	250	2.5	59.784	61.888	61.928	1.101	1.089	1.089
0.0100	1000	10	69.120	69.490	69.512	1.037	0.962	0.962
0.0100	1500	15	69.866	70.474	70.460	1.114	0.979	0.978
0.0100	2000	20	70.028	70.560	70.564	1.162	0.980	0.980

q	n	ΔL_{test}	$\Delta\mathrm{Acc}$
0.0050	500	0.154 ± 0.111	1.57 ± 2.10
0.0075	500	0.037 ± 0.055	0.04 ± 1.17
0.0100	500	0.026 ± 0.042	-0.03 ± 2.44
0.0200	500	-0.003 ± 0.027	-0.58 ± 1.03
0.0500	500	0.000 ± 0.005	-0.07 ± 0.20
0.0100	250	0.012 ± 0.022	2.10 ± 1.12
0.0100	1000	0.075 ± 0.063	0.37 ± 0.69
0.0100	1500	0.135 ± 0.084	0.61 ± 1.67
0.0100	2000	0.182 ± 0.080	0.53 ± 1.38

F Proofs for Section 2 and Appendix B

F.1 Proof of Theorem 2.1

Proof of Theorem 2.1. For arbitrary P and Q, take $\mu = P + Q$ and define

$$\ell(x) = \log \frac{dP/d\mu(x)}{dQ/d\mu(x)}.$$

Set $\ell = +\infty$ where only $dP/d\mu$ is positive and $\ell = -\infty$ where only $dQ/d\mu$ is positive. For $X \sim P$, the privacy loss $L = \ell(X)$ has no mass at $-\infty$; its mass at $+\infty$ is the mass of the part of P singular to Q. We use $e^{-\infty} = 0$. For $\gamma \geq 1$, write $H_\gamma(P\|Q) = \sup_A\{P(A) - \gamma Q(A)\}$.

Let $\mathcal{E}_{P,Q}$ be the closed convex testing-error region consisting of pairs

$$\big(\mathbb{E}_P[\phi], 1 - \mathbb{E}_Q[\phi]\big), \qquad 0 \leq \phi \leq 1.$$

The f-DP curve $T_{P,Q}$ is the lower boundary of $\mathcal{E}_{P,Q}$. For $\gamma \geq 1$, the two hockey-stick divergences are equivalently the support values

$$H_\gamma(P\|Q) = 1 - \inf_{(a,b)\in\mathcal{E}_{P,Q}} (a + \gamma b), \qquad H_\gamma(Q\|P) = 1 - \inf_{(a,b)\in\mathcal{E}_{P,Q}} (\gamma a + b).$$

As γ ranges over $[1, \infty)$, the two families cover all positive supporting slopes. Convex duality therefore shows that the pair of one-sided (ε, δ)-DP curves and the f-DP curve determine one another.

We next connect the two one-sided curves to privacy loss. The hockey-stick identity gives, for $x \geq 0$,

$$\delta_{P\|Q}(x) = \mathbb{E}_P\left[(1 - e^{x-L})_+\right], \qquad \delta'_{P\|Q,+}(x) = -e^x \mathbb{E}_P\left[e^{-L}\mathbf{1}_{\{L>x\}}\right].$$

Consequently,

$$P\{L > x\} = \delta_{P\|Q}(x) - \delta'_{P\|Q,+}(x), \qquad P\{L < -x\} = -e^{-x}\delta'_{Q\|P,+}(x).$$

These strict-tail identities, together with normalization and $P\{L = +\infty\} = \lim_{x\to\infty} \delta_{P\|Q}(x)$, recover the law of L under P. Conversely, that law determines both one-sided curves via

$$\delta_{P\|Q}(\varepsilon) = \mathbb{E}_P[(1 - e^{\varepsilon - L})_+], \qquad \delta_{Q\|P}(\varepsilon) = 1 - \mathbb{E}_P[e^{-L}] + \mathbb{E}_P[(e^{-L} - e^{\varepsilon})_+].$$

The identity $D_{1+t}(P\|Q) = K(t)/t$ then shows that the privacy-loss law determines the exact RDP curve.

For the converse, suppose $D_{\alpha_0}(P\|Q) < \infty$ for some $\alpha_0 > 1$. The RDP curve determines $K(t)$ on an interval $0 < t < c$. Fix $t_0 \in (0, c)$ and exponentially tilt the law of L by the density $e^{t_0 L - K(t_0)}$. The tilted law has moment-generating function

$$\mathbb{E}[e^{uL_{t_0}}] = \exp\{K(t_0 + u) - K(t_0)\}, \qquad -t_0 < u < c - t_0.$$

This function is known from RDP and is finite on a neighborhood of zero, so moment-generating-function uniqueness determines the tilted law. Untilting by

$$P\{L \in B\} = e^{K(t_0)}\mathbb{E}\Big[e^{-t_0 L_{t_0}}\mathbf{1}_{\{L_{t_0}\in B\}}\Big]$$

recovers the original privacy-loss law. The preceding formulas recover both one-sided curves, and convex duality for $\mathcal{E}_{P,Q}$ then recovers the f-DP curve. □

F.2 Proof of Corollary F.1

Corollary F.1 (Centrally symmetric additive noise)**.** *Let θ be a real-valued query and let $\mathcal{M}(D) = \theta(D) + Z$. Fix ordered neighboring datasets D, D', put $s = \theta(D) - \theta(D') \in \mathbb{R}$, and translate both outputs by $-\theta(D')$. Suppose $Z \overset{d}{=} -Z$, and set $P = \mathrm{Law}(s + Z)$, $Q = \mathrm{Law}(Z)$. Here s is the query difference for this fixed neighboring pair. The reflection $R_s(x) = s - x$ exchanges P and Q, so*

$$\delta_{P\|Q}(\varepsilon) = \delta_{Q\|P}(\varepsilon), \qquad D_\alpha(P\|Q) = D_\alpha(Q\|P), \qquad T_{P,Q} = T_{Q,P}.$$

The common one-sided (ε, δ)-DP curve, the exact f-DP curve, and the law of L determine one another and determine every exact RDP value through $D_{1+t}(P\|Q) = K(t)/t$. Conversely, if $D_{\alpha_0}(P\|Q) < \infty$ for some real $\alpha_0 > 1$, the exact RDP curve recovers the other three notions.

Proof of Corollary F.1. Set $R_s(x) = s - x$. If $X = s + Z \sim P$, then $R_s(X) = -Z \sim Q$ by central symmetry. If $X = Z \sim Q$, then $R_s(X) = s - Z \sim P$. Thus R_s exchanges P and Q, $R_s \circ R_s$ is the identity, and R_s preserves $\mu = P + Q$. By uniqueness of Radon–Nikodym derivatives, versions may be chosen so that $(dP/d\mu) \circ R_s = dQ/d\mu$ and $(dQ/d\mu) \circ R_s = dP/d\mu$ μ-almost everywhere. Consequently, $\ell \circ R_s = -\ell$, including when ℓ takes an extended value. Transforming measurable sets by R_s now gives $\delta_{P\|Q} = \delta_{Q\|P}$. Invariance of Rényi divergence under the bimeasurable transformation R_s similarly gives $D_\alpha(P\|Q) = D_\alpha(Q\|P)$. If a randomized test ϕ of P against Q has type I and type II errors (a, b), then $\phi \circ R_s$, viewed as a test of Q against P, also has error pair (a, b). Applying the reflection argument with (Q, P) in place of (P, Q) gives $T_{P,Q} = T_{Q,P}$.

We next derive the reflection identity for the privacy-loss distribution. On the common support, $dQ = e^{-\ell} dP$, and hence $Q\{\ell \in B\} = \mathbb{E}_P[e^{-L}\mathbf{1}_{\{L\in B\}}]$ for every Borel $B \subseteq \mathbb{R}$. Applying R_s and then this change of measure gives

$$P\{L \in -B\} = P\{-\ell \in B\} = P\{\ell \circ R_s \in B\} = Q\{\ell \in B\} = \mathbb{E}_{X\sim P}\left[e^{-L}\mathbf{1}_{\{L\in B\}}\right].$$

The displayed exponential-weighting identity determines the restriction of the law of L to $(-\infty, 0)$ from its restriction to $(0, \infty)$; the masses at zero and $+\infty$ are already part of the given restriction to $[0, +\infty]$, and a privacy-loss distribution under P has no atom at $-\infty$.

Because the two one-sided (ε, δ)-DP curves coincide, the common curve supplies both inputs required by Theorem 2.1. Theorem 2.1 gives the unconditional forward recoveries and, under the finite-order condition stated in Corollary F.1, the converse recovery from the RDP curve. □

F.3 Proof of Example B.1

Proof of Example B.1. The masses on the nonnegative and negative support points sum to $4c/3 = 8/11$ and $c/2 = 3/11$, respectively. Thus both are full-support probability mass functions. If $W_i(k) = P_i\{k\}/Q_i\{k\}$, direct division gives

	$k \geq 1$	$k = 0$	$k \leq -1$
$W_A(k)$	4	$1/4$	$1/2$
$W_B(k)$	4	$1/3$	$1/3$.

Since $H_\gamma(P_i\|Q_i) = \sum_k q_i(k)(W_i(k) - \gamma)_+$, only $k \geq 1$ contributes to the forward divergence when $\gamma \geq 1$. This gives the common formula in the example. Similarly, $H_\gamma(Q_i\|P_i) = \sum_k q_i(k)(1 - \gamma W_i(k))_+$ gives the two reverse formulas.

Writing F for the common forward one-sided (ε, δ)-DP curve and R_A, R_B for the reverse one-sided (ε, δ)-DP curves, direct subtraction gives

$$F - R_A = \begin{cases} c(\gamma - 1)/6, & 1 \leq \gamma \leq 2, \\ c(4 - \gamma)/12, & 2 \leq \gamma \leq 4, \end{cases}$$

and

$$F - R_B = \begin{cases} c(\gamma - 1)/6, & 1 \leq \gamma \leq 3, \\ c(4 - \gamma)/3, & 3 \leq \gamma \leq 4. \end{cases}$$

All curves vanish after their stated cutoffs, so the pointwise maxima agree. Finally,

$$\exp(D_2(P_i\|Q_i)) = \sum_k q_i(k) W_i(k)^2.$$

Substituting the table and the three region masses gives $265/88$ for pair A and 3 for pair B. The table also shows that both W_i and W_i^{-1} are bounded by 4. □

F.4 Proof of Example B.2

Proof of Example B.2. The masses of Q_w sum to one by the definition of $\mathrm{Li}_w(e^{-1})$. At each $k \geq 1$, the privacy loss is $L_w = k$, which gives the mean and the hockey-stick divergence stated in the example. For every $t > 0$,

$$\mathbb{E}_{X\sim P_w}\left[e^{tL_w}\right] = \frac{1}{\zeta(w)} \sum_{k=1}^{\infty} \frac{e^{tk}}{k^w} = +\infty.$$

Hence the RDP curve is everywhere infinite for both $w = 3$ and $w = 4$. □

G Proofs for Section 3 and Appendix C

G.1 Proof of Lemma G.1

Lemma G.1 (Hockey-stick Chernoff bound)**.** *For $x \in \mathbb{R}$ and $t > 0$,*

$$(1 - e^x)_+ \le \frac{t^t}{(1+t)^{1+t}} e^{-tx}.$$

Proof of Lemma G.1. If $x \ge 0$, the left side is zero. Now suppose $x < 0$ and set $y = e^x \in (0, 1)$. The function $y^t(1-y)$ is maximized at $y = t/(1+t)$, where $y^t(1-y)$ equals $t^t/(1+t)^{1+t}$. Therefore

$$(1 - e^x)e^{tx} \le \frac{t^t}{(1+t)^{1+t}},$$

which is the stated inequality. □

G.2 Proof of the conversion bounds in Appendix C.1

We prove the conversion bounds stated in Appendix C.1.

Applying Lemma G.1 with $x = \varepsilon - L$ and taking expectation under P gives

$$\delta_{P\|Q}(\varepsilon) \le \frac{t^t}{(1+t)^{1+t}} \exp\{K(t) - t\varepsilon\}.$$

Since $K(t) = tD_{t+1}(P\|Q)$, an order-$1+t$ bound $D_{t+1}(P\|Q) \le \gamma$ therefore implies

$$\delta_{P\|Q}(\varepsilon) \le \frac{t^t}{(1+t)^{1+t}} \exp\{t(\gamma - \varepsilon)\}.$$

This is the refined fixed-order conversion of Balle et al. (2020b); dropping the prefactor recovers the original conversion of Mironov (2017). If $\gamma = \widetilde{\rho}(1+t)$, optimization over $t > 0$ gives (C.6).

Let $b(t)$ denote the logarithm of the objective in (C.6). Its first two derivatives are

$$b'(t) = \log \frac{t}{1+t} + \widetilde{\rho}(2t+1) - \varepsilon, \qquad b''(t) = \frac{1}{t(1+t)} + 2\widetilde{\rho} > 0.$$

For $\widetilde{\rho} > 0$, the first derivative tends to $-\infty$ as $t \downarrow 0$ and to $+\infty$ as $t \to \infty$. Hence the minimizer is unique and satisfies (C.7). Finally, when $\rho > 0$, substitute $K(t) = \rho t(t+1) - \Gamma(t)$ into (C.1) at $t = t_\varepsilon(\rho)$ to obtain (C.8).

G.3 Proof of Proposition 3.2

Proof of Proposition 3.2. Since $F_{\mathrm{R}}(t_{\mathrm{Z}}) \ge \varepsilon_{\mathrm{R}}$,

$$\varepsilon_{\mathrm{Z}} = F_{\mathrm{R}}(t_{\mathrm{Z}}) + \frac{\Gamma(t_{\mathrm{Z}})}{t_{\mathrm{Z}}} \ge \varepsilon_{\mathrm{R}} + \frac{\Gamma(t_{\mathrm{Z}})}{t_{\mathrm{Z}}}.$$

Similarly, evaluating F_{Z} at t_{R} gives

$$\varepsilon_{\mathrm{Z}} \le F_{\mathrm{Z}}(t_{\mathrm{R}}) = \varepsilon_{\mathrm{R}} + \frac{\Gamma(t_{\mathrm{R}})}{t_{\mathrm{R}}}.$$

The final assertion follows from $\Gamma(t) \ge 0$. □

G.4 Proof of Lemma G.2

Lemma G.2 isolates the quantities needed for a quantitative fourth-order expansion: the excess of $\rho(s)$ over $\mathbb{E}_{P_s} L_s$ and the third and fourth cumulants of privacy loss. The lemma applies equally to continuous and discrete mechanisms and retains an explicit fifth-order remainder.

Lemma G.2 (Quantitative fourth-order identity for the CGF gap)**.** *Fix $T > 0$. Let L_s be the privacy-loss variables of mutually absolutely continuous pairs, with CGF K, so that $K(-1) = 0$. As usual, the dependence of K on s is implicit. Suppose, as $s \to 0$ through nonzero values, that $\rho(s) < \infty$, $K \in C^5([-1, T])$, and there is an $M_T < \infty$ such that*

$$\sup_{-1 \le u \le T} |K^{(5)}(u)| \le M_T |s|^5. \tag{G.1}$$

Then, for $0 < t \le T$,

$$\rho(s)t(t+1) - K(t) = (\rho(s) - \kappa_1(L_s))t(t+1) - \frac{\kappa_3(L_s)}{6}t^2(t+1) + \frac{\kappa_4(L_s)}{24}t^2(1-t^2) + E_s(t).$$

Moreover,

$$|E_s(t)| \le \frac{M_T|s|^5}{120}(t^2 + t^5), \qquad 0 < t \le T. \tag{G.2}$$

Proof of Lemma G.2. Taylor's theorem and (G.1) give

$$K(u) = \sum_{j=1}^{4} \frac{\kappa_j(L_s)}{j!}u^j + \Delta_s(u), \qquad |\Delta_s(u)| \le \frac{M_T|s|^5}{120}|u|^5$$

for $-1 \le u \le T$. Applying this identity at $u = -1$ gives

$$\kappa_1(L_s) = \frac{\kappa_2(L_s)}{2} - \frac{\kappa_3(L_s)}{6} + \frac{\kappa_4(L_s)}{24} + \Delta_s(-1).$$

Substituting this identity into the cumulant expansion of $K(t)$ gives

$$\kappa_1(L_s)t(t+1) - K(t) = -\frac{\kappa_3(L_s)}{6}t^2(t+1) + \frac{\kappa_4(L_s)}{24}t^2(1-t^2) + t^2\Delta_s(-1) - \Delta_s(t).$$

Thus $E_s(t) = t^2\Delta_s(-1) - \Delta_s(t)$. The two Taylor-remainder bounds give (G.2); adding $(\rho(s) - \kappa_1(L_s))t(t+1)$ proves the identity. □

G.5 Proof of Theorem 3.1

Proof of the local expansion theorem, Theorem 3.1. Apply Lemma G.2 with $\kappa_1(L_s) = \mathbb{E}_{P_s} L_s$. Its identity is (3.3), and (G.2) gives (3.4). □

G.6 Proof of Theorem 3.3

Proof of Theorem 3.3. Independence gives

$$tD_{t+1}(P^{(n)} \| Q^{(n)}) = \sum_{i=1}^{n} K_i(t) = \boldsymbol{K}(t).$$

Subtracting this identity from $\rho_{\mathrm{sum}}t(t+1) = \sum_i \rho_i t(t+1)$ proves (C.10). For n copies of one pair, $\boldsymbol{K} = nK$, so $\sup_{t>0} \boldsymbol{K}(t)/\{t(t+1)\} = n\rho$ and $\boldsymbol{\Gamma} = n\Gamma$. □

G.7 Proof of (C.11) and (C.12)

Under the additional local-expansion assumptions in Appendix C.2, applying the local expansion theorem, Theorem 3.1, to the components $i \in \mathcal{I}$ and observing that every component outside $\mathcal{I}$ contributes zero proves (C.11). Finally, the triangle inequality and (3.4) give (C.12).

H PROOFS FOR SECTION 4 AND APPENDIX D

H.1 PROOF OF LEMMA H.1

Throughout the additive-noise calculations, s is a signed output shift. Symmetric location families are reduced to the magnitude $r = |s|$ when a positive RDP order must be rescaled.

For the following endpoint-localization lemma, specialize the location convention (3.2) to a noise variable Z with density f. It verifies $\rho(s) = K'(0) = \mathbb{E}_{P_s} L_s$ for Student-t noise. The privacy-loss and CGF definitions give

$$L_s = \log f(Z) - \log f(Z+s), \qquad K(t) = \log \mathbb{E}_{Z\sim f}[e^{tL_s}],$$

and define the location score and its CGF

$$S := -(\log f)'(Z), \qquad \Lambda_S(\tau) := \log \mathbb{E}_{Z\sim f}[e^{\tau S}].$$

These equalities hold wherever the expectations are finite. Since $\mathbb{E}_{Z\sim f}[S] = 0$, the fourth cumulant used below is

$$\kappa_4(S) = \mathbb{E}_{Z\sim f}[S^4] - 3\big(\mathbb{E}_{Z\sim f}[S^2]\big)^2.$$

All remainders in the following lemma are uniform over each fixed compact interval of positive RDP-order offsets.

Lemma H.1 (Uniform control over large RDP orders)**.** *Suppose f is even and $P_s \sim Q$ for all sufficiently small nonzero s. Put $r = |s|$. Symmetry makes the CGFs at shifts s and r identical; for the current shift define*

$$\mathcal{H}_s(\tau) = \frac{K(\tau/r)}{\tau(\tau+r)}, \qquad \tau > 0.$$

Assume that $\mathcal{H}_s$ extends continuously to $\tau = 0$ with $\mathcal{H}_s(0) = K'(0)/r^2$, and that

$$\frac{K'(0)}{s^2} \longrightarrow \frac{1}{2}\mathbb{E}_{Z\sim f}[S^2], \qquad K(t) = \frac{1}{2}\mathbb{E}_{Z\sim f}[S^2]s^2 t(t+1) + o(|s|^2) \tag{H.1}$$

uniformly on every compact $J \subset (0,\infty)$.

For endpoint localization, assume that differentiation under the defining integral is valid through the required order and that the following uniform bound holds for every sufficiently small fixed $\eta > 0$:

$$\limsup_{s\to 0} \sup_{\tau\geq\eta} \mathcal{H}_s(\tau) \leq \sup_{\tau\geq\eta} \frac{\Lambda_S(\tau)}{\tau^2}. \tag{H.2}$$

Near $(r,\tau) = (0,0)$ suppose the Taylor remainder is uniform and satisfies

$$\mathcal{H}_s(\tau) - \mathcal{H}_s(0) = \frac{\kappa_4(S)}{24}\tau(\tau+r) + O\big(\tau(\tau+r)^2\big). \tag{H.3}$$

Assume $\kappa_4(S) < 0$ and, for some $a_4 > 0$,

$$\frac{1}{2}\mathbb{E}_{Z\sim f}[S^2] - \frac{\Lambda_S(\tau)}{\tau^2} \geq a_4 \min(\tau^2, 1), \qquad \tau > 0. \tag{H.4}$$

Then

$$\rho(s) = K'(0) = \mathbb{E}_{P_s} L_s$$

for all sufficiently small nonzero s. Hence $A_s = 0$ exactly. The remainders in (H.1) *are uniform for t in a fixed compact $J \subset (0,\infty)$, and the remainder in* (H.3) *is uniform in the stated neighborhood of $(r,\tau) = (0,0)$.*

Proof of Lemma H.1. We begin with the factorization used for endpoint localization. The likelihood identity gives $K(0) = K(-1) = 0$. Therefore, the definition of $\mathcal{H}_s$ gives the exact factorization

$$K(t) = r^2 t(t+1)\mathcal{H}_s(rt).$$

Moreover, $K'(0) = \mathbb{E}_{P_s} L_s$. Dividing the numerator and denominator in the definition of $\mathcal{H}_s(\tau)$ by τ and letting $\tau \to 0$ gives $\mathcal{H}_s(0) = K'(0)/r^2$, as stated.

Choose $\eta > 0$ small enough that, for all $|s|$ sufficiently small, the absolute value of the remainder in (H.3) is bounded by

$$-\frac{\kappa_4(S)}{48}\tau(\tau + r)$$

whenever $0 < \tau \le \eta$. It follows that

$$\mathcal{H}_s(\tau) - \mathcal{H}_s(0) \le \frac{\kappa_4(S)}{48}\tau(\tau + r) < 0, \qquad 0 < \tau \le \eta.$$

For $\tau \ge \eta$, (H.2) and (H.4) put the supremum a fixed distance below $\frac{1}{2}\mathbb{E}_{Z\sim f}[S^2]$. Because $\mathcal{H}_s(0) = K'(0)/r^2$ tends to $\frac{1}{2}\mathbb{E}_{Z\sim f}[S^2]$, the region $\tau \ge \eta$ also lies below $\mathcal{H}_s(0)$ for all sufficiently small nonzero s. Hence $r^2 \sup_{\tau>0} \mathcal{H}_s(\tau) = K'(0)$, which gives $\rho(s) = K'(0) = \mathbb{E}_{P_s} L_s$.

This completes the proof. □

The hypotheses of Lemma H.1 can be checked from K and Λ_S without first evaluating ρ. For a C^5 log-density, a common integrable bound on the required derivatives justifies the Taylor expansions in (H.3). The uniform bound in (H.2) must be verified separately for each mechanism, as for the Student-t_ν family in Proposition D.2. Nondifferentiable log-densities require piecewise calculations. The sign and uniform-control conditions exclude a zero fourth cumulant and rule out a competing second-order maximum away from the origin.

H.2 Proof of (H.5) and (H.6)

For a smooth location family, continue to use the location-score notation S introduced above. Suppose $\log f$ is five times continuously differentiable, differentiation under the relevant expectations is dominated through fifth order, the required score moments are finite, and the integration-by-parts boundary terms vanish. Taylor expansion of $L_s = \log f(Z) - \log f(Z + s)$ and integration by parts then give

$$\kappa_3(L_s) = \kappa_3(S)s^3 + \frac{1}{2}\kappa_4(S)s^4 + O(|s|^5), \tag{H.5}$$

$$\kappa_4(L_s) = \kappa_4(S)s^4 + O(|s|^5). \tag{H.6}$$

Indeed, the integration-by-parts identities

$$\mathbb{E}[(\log f)''(Z)] = -\mathbb{E}[((\log f)'(Z))^2],$$

$$\mathbb{E}[((\log f)'(Z))^2(\log f)''(Z)] = -\frac{1}{3}\mathbb{E}[((\log f)'(Z))^4]$$

give the fourth-order coefficient in (H.5). These identities also show why a kink in the log-density can change the fourth-order coefficient.

H.3 Proof of Proposition 4.1

Exact CGFs. In the setting of Proposition 4.1, the following hold for every $s \in \mathbb{R}$ and $t > 0$. For Gaussian noise,

$$K(t) = \frac{t(t+1)}{2\sigma^2}s^2. \tag{H.7}$$

For the Gaussian mixture, write $Z = m\xi + G$ with ξ uniform on $\{-1, 1\}$ and $G \sim N(0, 1)$ independent; then

$$K(t) = \log \mathbb{E}\left[\exp\{t(sZ + s^2/2)\}\left\{\frac{\cosh(mZ)}{\cosh(m(Z+s))}\right\}^t\right]. \tag{H.8}$$

For unit-scale Laplace noise,

$$K(t) = \log\left(\frac{(1+t)e^{t|s|} + te^{-(1+t)|s|}}{1+2t}\right). \tag{H.9}$$

Proof of Proposition 4.1. For part (a), completing the square in the Gaussian likelihood ratio gives (H.7) and the stated optimal parameter. Under P_s,

$$L_s = \frac{sZ}{\sigma^2} + \frac{s^2}{2\sigma^2}, \qquad Z \sim N(0, \sigma^2),$$

so $A_s = 0$, $U_s = Z/\sigma^2$, and $\kappa_3(U_s) = \kappa_4(U_s) = 0$. The CGF is quadratic, so the remainder in the local expansion theorem also vanishes. The theorem therefore gives $\Gamma(t) = 0$ exactly.

For part (b), the mixture density is proportional to $e^{-x^2/2}\cosh(mx)$, so

$$(\log f_m)'(x) = -x + m\tanh(mx), \qquad (\log f_m)''(x) = -1 + m^2\operatorname{sech}^2(mx).$$

Evaluating $\log\{f_m(Z)/f_m(Z+s)\}$ gives (H.8). Integration by parts gives

$$\mathbb{E}[((\log f_m)'(Z))^2] = 1 - m^2\mathbb{E}[\operatorname{sech}^2(mZ)].$$

Generating $Z + s$ by first shifting a unit-variance Gaussian and then adding the independent term $m\xi$ shows by data processing that $\rho(s) \le s^2/2$. The density is even, so both the CGF and $\rho(s)$ are even in s. For $z \ge 0$,

$$\log \frac{f_m(z)}{f_m(z+|s|)} \ge |s|z + \frac{s^2}{2} - m|s|, \qquad f_m(z) \ge \frac{1}{2}\phi(z-m).$$

Consequently,

$$e^{K(t)} \ge \frac{1}{2} e^{s^2 t(t+1)/2}\mathbb{P}_{Y\sim N(m+t|s|,1)}[Y \ge 0].$$

After division by $t(t+1)$ and passage to $t \to \infty$, this proves the matching lower bound $\rho(s) \ge s^2/2$.

Put

$$S_m := -(\log f_m)'(Z) = Z - m\tanh(mZ), \qquad I_m := \mathbb{E}[S_m^2].$$

The integration-by-parts identity above gives $I_m = 1 - m^2\mathbb{E}[\operatorname{sech}^2(mZ)]$. Since f_m is even, $\mathbb{E}_{P_s} L_s$ is even in s, and its second-order expansion is

$$\mathbb{E}_{P_s} L_s = \frac{I_m}{2} s^2 + O(s^4).$$

Together with $\rho(s) = s^2/2$, this yields

$$A_s = m^2\mathbb{E}[\operatorname{sech}^2(mZ)] + O(s^2).$$

Moreover, S_m is symmetric. The smooth-location cumulant expansions (H.5)–(H.6) give

$$\kappa_3(U_s) = \tfrac{1}{2}\kappa_4(S_m)s + O(s^2) = O(s), \qquad \kappa_4(U_s) = \kappa_4(S_m) + O(|s|) = O(1).$$

Substitution in the local expansion theorem, Theorem 3.1, proves (4.1).

For part (c), symmetry makes the CGF depend on the shift only through $|s|$. Splitting at the two kinks gives (H.9). The exact value $\rho(s) = |s| + e^{-|s|} - 1$ is the continuous-Laplace formula established by Harrison & Manurangsi (2026); differentiating the exact CGF at zero also gives $\mathbb{E}_{P_s} L_s = \rho(s)$ and hence $A_s = 0$. For $s > 0$, differentiation of the same transform gives

$$\kappa_3(U_s) = -s + O(s^2), \qquad \kappa_4(U_s) = -2 + O(s).$$

Symmetry gives $U_{-s} \overset{d}{=} -U_s$, so the signed-$s$ form of the first display is $\kappa_3(U_s) = -s + O(s^2)$, while the fourth-cumulant remainder is $O(|s|)$. The local expansion theorem therefore gives

$$\begin{aligned}\Gamma(t) &= \frac{s^4}{6}t^2(t+1) - \frac{s^4}{12}t^2(1-t^2) + O(|s|^5) \\ &= \frac{[t(t+1)]^2}{12}s^4 + O(|s|^5),\end{aligned}$$

which proves item *(c)*.

It remains to verify the common remainder hypothesis used above. It is immediate for Gaussian noise. For the Gaussian mixture and $|s| \le s_0$,

$$|L_s| \le |s|\big(|Z| + m + s_0/2\big),$$

and the mixture has uniformly bounded exponentially tilted moments of every fixed order. For Laplace noise,

$$|L_s| \le |s|.$$

Thus the fifth absolute centered moments under every tilt with $-1 \le u \le T$ are $O(|s|^5)$. Since $K^{(5)}(u)$ is the fifth cumulant under the tilt, it is also $O(|s|^5)$ uniformly on this interval. The apparent singularities in the displayed Laplace formulas are removable because these bounded privacy-loss variables have positive smooth moment-generating functions on the real line. Hence all three families satisfy the hypotheses of the local expansion theorem.

□

H.4 Proof of Proposition 4.2

Proof of Proposition 4.2. Fix $z > 0$, let $Q = N(0, z^2C^2I)$, and, for $0 \le q \le 1$, set $P_q = (1-q)Q + qN(\mathbf{v}, z^2C^2I)$. Let W be the likelihood ratio of $N(\mathbf{v}, z^2C^2I)$ with respect to Q. Under Q,

$$\log W = \frac{G}{z} - \frac{1}{2z^2}, \qquad G \sim N(0,1),$$

and hence, for $\alpha \ge 1$,

$$\mathbb{E}_Q[W^\alpha] = \exp\left\{\frac{\alpha(\alpha-1)}{2z^2}\right\}. \tag{H.10}$$

Put $W_q = 1 - q + qW = dP_q/dQ$. Convexity of $x \mapsto x^\alpha$ gives

$$\mathbb{E}_Q[W_q^\alpha] \le 1 - q + q\mathbb{E}_Q[W^\alpha] \le \exp\left\{\frac{\alpha(\alpha-1)}{2z^2}\right\}.$$

Thus the one-step optimal zCDP parameter is at most $1/(2z^2)$. Since $W_q \ge qW$,

$$K(t) \ge (t+1)\log q + \frac{t(t+1)}{2z^2}.$$

Dividing by $t(t+1)$ and letting $t \to \infty$ proves $\rho(q) = 1/(2z^2)$ for every $q > 0$, which proves the one-step identities in (4.2).

At $t = 1$,

$$e^{K(1)} = \mathbb{E}_Q[W_q^2] = 1 + q^2\mathbb{E}_Q[(W-1)^2] = 1 + \left(e^{1/z^2} - 1\right)q^2,$$

which proves the formula for $\Gamma(1)$.

It remains to establish monotonicity. For fixed z and $t > 0$, write

$$M_t(q) := \mathbb{E}_Q\left[(1 + q(W-1))^{t+1}\right].$$

Then $M_t'(0) = (t+1)\mathbb{E}(W-1) = 0$, whereas, for $0 < q < 1$,

$$M_t''(q) = t(t+1)\mathbb{E}_Q\left[(W-1)^2\{1 + q(W-1)\}^{t-1}\right] > 0.$$

Hence $M_t(q)$, and therefore $K(t)$ viewed as a function of q, is strictly increasing. The one-step gap is consequently strictly decreasing. At $q = 1$, the pair is an ordinary Gaussian shift and its gap is zero. Continuity of the expectation at $q = 0$ gives the displayed one-step right-hand limit.

For part (b), set $z = z_n$, so that $1/(2z_n^2) = \rho_{\mathrm{sum}}/n$. Exact product additivity gives $\boldsymbol{K}(t) = nK(t)$, and its normalized supremum is $n/(2z_n^2) = \rho_{\mathrm{sum}}$. Thus the standard-composition parameter is optimal for the product pair, and item *(b)* follows. The monotonicity and endpoint statements follow immediately from part (a). □

H.5 Proof of Proposition D.1

Proof of Proposition D.1. Independence makes the product likelihood ratio the product of the component likelihood ratios, so their log moment-generating functions add. Hence $\boldsymbol{\Gamma}(t) = \sum_i \Phi_t(\rho_i)$ exactly. If Φ_t is convex, Jensen's inequality and $\sum_i \rho_i = \rho_{\mathrm{sum}}$ give

$$\boldsymbol{\Gamma}(t) \ge n\Phi_t\left(\frac{\rho_{\mathrm{sum}}}{n}\right).$$

Because $\psi(0) = 0$ and the single-pair gap is zero there, the chord inequality between 0 and ρ_{sum} gives

$$\Phi_t(\rho_i) \le \frac{\rho_i}{\rho_{\mathrm{sum}}} \Phi_t(\rho_{\mathrm{sum}}).$$

Summing this inequality proves the upper bound in (D.6). Concavity reverses Jensen's inequality and the chord inequality, and therefore reverses both bounds. The equality statements follow from the strict forms of the same two inequalities.

For the unit-scale Laplace illustration, parameterize the selected inverse branch by the magnitude $r \ge 0$. Here $x = \rho(r) = r + e^{-r} - 1$, equivalently $r = \psi(x)$. Direct differentiation of the exact expressions in Proposition 4.1 gives, for $r > 0$,

$$\frac{d^2}{dx^2} \Phi_1(x) \Big|_{x=\rho(r)} = \frac{2e^{2r}\{2e^{3r} + 6e^{2r} + 3e^r + 1\}}{(2e^{3r} + 1)^2} > 0.$$

The right limit at $r = 0$ is $8/3$, so Φ_1 is strictly convex throughout the interval used in Figure 7. For the Gaussian-mixture panel, $\psi(x) = \sqrt{2x}$, and strict concavity of Φ_1 on the displayed interval is suggested by centered finite differences on the plotted grid; this observation concerns the displayed numerical grid. □

H.6 Proof of Proposition D.2

Proof of Proposition D.2. Fix ν and retain the notation Z, P_s, Q and $\rho_\nu(s)$ from the proposition. Write $\ell_\nu = \log f_\nu$. The location score is

$$S_\nu := -\ell_\nu'(Z) = \frac{(\nu+1)Z}{\nu + Z^2}, \qquad |S_\nu| \le B_\nu := \frac{\nu+1}{2\sqrt{\nu}}.$$

The density is even, so the CGF and $\rho_\nu(s)$ are even in s. Put $r = |s|$. The mean-value theorem gives $|L_s| \le B_\nu r$. Consequently,

$$0 \le \frac{K(t)}{t(t+1)} \le \frac{B_\nu r}{t+1},$$

which tends to zero as $t \to \infty$. The normalized CGF is continuous on compact t-intervals and has the finite endpoint $K'(0) = \mathbb{E}_{P_s} L_s$, so $\rho_\nu(s) < \infty$ for every $s \in \mathbb{R}$.

To localize the envelope for all sufficiently small nonzero s, put $Y_\nu = Z^2/(\nu + Z^2)$. The beta representation of a Student random variable gives $Y_\nu \sim \mathrm{Beta}(1/2, \nu/2)$ and hence, for $j \ge 1$,

$$\mu_{2j,\nu} := \mathbb{E}[S_\nu^{2j}] = \frac{(\nu+1)^{2j}}{\nu^j} \frac{(1/2)_j (\nu/2)_j}{((\nu+1)/2)_{2j}}. \tag{H.11}$$

In particular,

$$I_\nu := \mu_{2,\nu} = \frac{\nu+1}{\nu+3}, \qquad \kappa_4(S_\nu) = -\frac{18(\nu+1)^2(\nu^2+4\nu-1)}{\nu(\nu+3)^2(\nu+5)(\nu+7)} < 0. \tag{H.12}$$

Comparison with the even moments of $N(0, I_\nu)$ gives

$$\frac{\mu_{2j+2,\nu}/\mu_{2j,\nu}}{(2j+1)I_\nu} = \frac{(\nu+1)(\nu+3)(2j+\nu)}{\nu(4j+\nu+1)(4j+\nu+3)} < 1, \qquad j \ge 1,$$

where the difference between denominator and numerator is $2j(8\nu j + 3\nu^2 + 4\nu - 3) > 0$ for $\nu \ge 1$. Because S_ν is symmetric and bounded, termwise comparison of the entire moment series yields

$$\mathbb{E}[e^{\tau S_\nu}] < e^{I_\nu \tau^2/2}, \qquad \tau \ne 0.$$

Equation (H.12) supplies a quadratic separation near $\tau = 0$, continuity supplies uniform separation on compact intervals away from zero, and $\log \mathbb{E}[e^{\tau S_\nu}] \le B_\nu \tau$ controls large τ. Thus the quantitative sub-Gaussian condition (H.4) holds.

All derivatives of positive order of ℓ_ν are bounded. Uniformly in Z,

$$L_r = -\sum_{j=1}^{5} \frac{r^j}{j!} \ell_\nu^{(j)}(Z) + O(r^6),$$

so $L_r/r \to S_\nu$ uniformly. Expanding the log moment-generating function and using the usual integration-by-parts identities gives (H.1) and (H.3), uniformly on the stated compact sets. Symmetry makes the CGFs at shifts s and $r = |s|$ equal, and $K(\tau/r) \to \log \mathbb{E}[e^{\tau S_\nu}]$ locally uniformly in τ. The bound

$$0 \le \frac{K(\tau/r)}{\tau(\tau + r)} \le \frac{B_\nu}{\tau + r}$$

controls $\tau \ge R$ uniformly once R is large, while local uniform convergence controls each interval $[\eta, R]$. This proves (H.2). Lemma H.1 therefore implies

$$\rho_\nu(s) = \mathbb{E}_{P_s} L_s$$

for every sufficiently small nonzero s.

Thus $A_s = 0$. Symmetry gives $\kappa_3(S_\nu) = 0$, and (H.5)–(H.6), followed by the local expansion theorem, Theorem 3.1, gives

$$\Gamma(t) = -\frac{\kappa_4(S_\nu)}{24}[t(t+1)]^2 s^4 + O(|s|^5).$$

The coefficient obtained from (H.12) is exactly $c_{4,\nu}$ in (D.8). Bounded derivatives justify six uniform differentiations in s. Since both $K(t)$ and $\mathbb{E}_{P_s} L_s = \rho_\nu(s)$ are even functions of s, the gap is even and its fifth-order Taylor coefficient vanishes. The remainder is therefore $O(|s|^6)$.

Finally,

$$c'_{4,\nu} = -\frac{3(\nu+1)p(\nu)}{4\nu^2(\nu+3)^3(\nu+5)^2(\nu+7)^2},$$

where

$$p(\nu) = \nu^6 + 8\nu^5 - 19\nu^4 - 288\nu^3 - 613\nu^2 - 72\nu - 105.$$

The coefficients of p have one sign change, so Descartes' rule shows that p has a unique positive root. Since $p(1) < 0$ and $p(\nu) \to \infty$, this root, numerically $6.013052641\ldots$, is the unique maximizer on $[1, \infty)$. Direct expansion gives $c_{4,\nu} = 3/(4\nu) - 9/\nu^2 + O(\nu^{-3})$. □

H.7 Proof of Proposition D.3

Proof of Proposition D.3. The density f_β is even, so the CGF and $\rho_\beta(s)$ are even functions of the shift, and $U_{-s} \overset{d}{=} -U_s$. Put $r = |s|$ when only the shift magnitude is relevant. Under P_s, write

$$L_s = \frac{|Z+s|^\beta - |Z|^\beta}{\beta}.$$

For $1 < \beta < 2$, the mean-value inequality and Young's inequality give, for $t \ge 1$,

$$tL_s \le tr(|Z| + r)^{\beta-1} \le \frac{|Z|^\beta}{2\beta} + C_{\beta,r}(1 + t^\beta).$$

Integration against f_β therefore yields $K(t) \le C'_{\beta,r}(1 + t^\beta)$, so $K(t)/\{t(t+1)\} \to 0$ as $t \to \infty$. Continuity on compact order intervals and the finite endpoint $K'(0) = \mathbb{E}_{P_s} L_s$ prove $\rho_\beta(s) < \infty$. For $\beta = 1$, $|L_s| \le r$ gives the same conclusion. For $\beta = 2$, completing the square gives $K(t) = s^2 t(t+1)/2$.

The absolute moments of Z are

$$\mathbb{E}|Z|^p = \beta^{p/\beta}\frac{\Gamma((p+1)/\beta)}{\Gamma(1/\beta)}, \qquad p > -1.$$

For this proof, define

$$S_\beta(z) := \operatorname{sgn}(z)|z|^{\beta-1}, \qquad I_\beta := \mathbb{E}[S_\beta(Z)^2], \qquad J_\beta := \mathbb{E}[S_\beta(Z)^4].$$

Below, an unqualified S_β or S'_β inside an expectation is evaluated at Z. The moment formula gives

$$I_\beta = \beta^{2-2/\beta}\frac{\Gamma(2 - 1/\beta)}{\Gamma(1/\beta)}, \qquad J_\beta = \beta^{4-4/\beta}\frac{\Gamma(4 - 3/\beta)}{\Gamma(1/\beta)}.$$

In particular, $J_\beta - 3I_\beta^2 = -24\chi_\beta$. To prove the sign of χ_β, put $x = 1/\beta \in [1/2, 1]$ and

$$\mathcal{C}(x) := \frac{J_\beta}{I_\beta^2} = \frac{\Gamma(4-3x)\Gamma(x)}{\Gamma(2-x)^2}.$$

The integral representation of the digamma function gives

$$\frac{d}{dx}\log \mathcal{C}(x) = \int_0^\infty \frac{3e^{-(4-3x)u} - e^{-xu} - 2e^{-(2-x)u}}{1-e^{-u}}\, du.$$

For $x < 1$, the numerator equals

$$e^{-xu}(y^2-1)(3y^2+1) < 0, \qquad y = e^{-(1-x)u} \in (0,1).$$

Thus $\mathcal{C}$ decreases from $\mathcal{C}(1/2) = 3$ to $\mathcal{C}(1) = 1$, proving $J_\beta \le 3I_\beta^2$ and hence $\chi_\beta \ge 0$.

Define the difference quotient

$$V_s(z) := \frac{|z+s|^\beta - |z|^\beta}{\beta s}.$$

Then $L_s = sV_s(Z)$ and, by (C.9),

$$U_s = V_s(Z) - \mathbb{E}[V_s(Z)],$$

and $V_s(Z) \to S_\beta(Z)$ in every fixed L^p. Consequently,

$$\kappa_4(U_s) = J_\beta - 3I_\beta^2 + o(1).$$

For $1 < \beta < 2$, write $S'_\beta(z) = (\beta-1)|z|^{\beta-2}$ for $z \neq 0$. For $k = 1, 2, 3$, split the integral defining $\mathbb{E}[V_s^k]$ into $|z| \le 2|s|$ and $|z| > 2|s|$. On the first region, both V_s and S_β are $O(|s|^{\beta-1})$, so the absolute value of its contribution to $s^{-1}\{\mathbb{E}[V_s^k] - \mathbb{E}[S_\beta^k]\}$ is $O(|s|^{k(\beta-1)}) = o(1)$. On the second region, Taylor's formula gives

$$V_s(z) = S_\beta(z) + \frac{s}{2}S'_\beta(z) + o\big(|s|\,|z|^{\beta-2}\big).$$

The required dominating powers are integrable because, for $k \le 3$, $S_\beta^{k-1}S'_\beta$ behaves at zero as $|z|^{(k-1)(\beta-1)+\beta-2}$, whose exponent is greater than -1. Hence

$$\frac{d}{ds}\mathbb{E}[V_s^k]\bigg|_{s=0} = \frac{k}{2}\mathbb{E}[S_\beta^{k-1}S'_\beta], \qquad k = 1, 2, 3.$$

Since $\mathbb{E}[S_\beta] = \mathbb{E}[S_\beta^3] = 0$, differentiating $\kappa_3(X) = \mathbb{E}X^3 - 3(\mathbb{E}X)(\mathbb{E}X^2) + 2(\mathbb{E}X)^3$ gives

$$\lim_{s\to 0}\frac{\kappa_3(U_s)}{s} = \frac{3}{2}\{\mathbb{E}[S_\beta^2 S'_\beta] - I_\beta \mathbb{E}[S'_\beta]\}.$$

Integration by parts yields $\mathbb{E}[S'_\beta] = I_\beta$ and $\mathbb{E}[S_\beta^2 S'_\beta] = J_\beta/3$. Hence

$$\kappa_3(U_s) = -12\chi_\beta s + o(|s|), \qquad \kappa_4(U_s) = -24\chi_\beta + o(1). \tag{H.13}$$

For $\beta = 1$, differentiation of the exact Laplace transform gives $\kappa_3(U_s) = -s + O(s^2)$ and $\kappa_4(U_s) = -2 + O(|s|)$, the same identity. At $\beta = 2$ both cumulants vanish.

It remains to verify the fifth-derivative hypothesis. For the pair (P_s, Q), let $P_{s,u}$ have density $dP_{s,u}/dP_s = \exp\{uL_s - K(u)\}$. Then $K^{(5)}(u)$ is the fifth cumulant of L_s under $P_{s,u}$. For $0 < |s| \le 1$,

$$|L_s| \le |s| G_\beta(Z), \qquad G_\beta(z) := (|z|+1)^{\beta-1}.$$

Let $C_T = \max(1, T)$. Since $\beta - 1 < \beta$,

$$\int (1 + G_\beta(z)^5) e^{C_T G_\beta(z)} f_\beta(z)\, dz < \infty.$$

Moreover $\sup_{-1\le u\le T}|K(u)| \to 0$, so the normalizing constants of these tilted laws are bounded away from zero for all sufficiently small nonzero s. Consequently,

$$\sup_{-1\le u\le T}\mathbb{E}_{P_{s,u}}|L_s - \mathbb{E}_{P_{s,u}}L_s|^5 \le C_{\beta,T}|s|^5.$$

The fifth cumulant equals the fifth centered moment minus ten times the product of the third and second centered moments; Lyapunov's inequality bounds that product by the fifth absolute centered moment. Therefore $\sup_{-1\le u\le T}|K^{(5)}(u)| \le C'_{\beta,T}|s|^5$. For $\beta = 1$, the same conclusion follows from $|L_s| \le |s|$; for $\beta = 2$ it is immediate from the quadratic CGF.

Insert (H.13) in the local expansion theorem, Theorem 3.1. The cubic and quartic terms combine to $\chi_\beta[t(t+1)]^2 s^4 + o(|s|^4)$. Since

$$\frac{t(t+1)}{2}\,\frac{2\{\rho_\beta(s) - \mathbb{E}_{P_s} L_s\}}{s^2}\,s^2 = t(t+1)\{\rho_\beta(s) - \mathbb{E}_{P_s} L_s\},$$

this proves (D.9).

The endpoint $t \downarrow 0$ in the definition of $\rho_\beta(s)$ gives $\rho_\beta(s) \ge \mathbb{E}_{P_s} L_s$. Replacing $\rho_\beta(s)$ by this smaller endpoint value shows that the first term on the right-hand side of (D.9) is nonnegative. This proves (D.10).

At $\beta = 1$, the exact Laplace transform in Proposition 4.1 gives $K'(0) = |s| + e^{-|s|} - 1 = \rho_1(s)$, and $I_1 = J_1 = 1$ gives $\chi_1 = 1/12$. At $\beta = 2$, the quadratic Gaussian CGF gives $\rho_2(s) = \mathbb{E}_{P_s} L_s = s^2/2$, $\chi_2 = 0$, and $\Gamma(t) = 0$ for every $t > 0$. □

H.8 Proof of Proposition D.4

Proof of Proposition D.4. For part (a), the discrete Laplace privacy loss is ar for $z \ge 0$, $-ar$ for $z \le -r$, and $a(r-2j)$ at $z = -j$ for $1 \le j < r$. Therefore

$$e^{K(t)} = \frac{1-\lambda}{1+\lambda}\left\{ e^{tar}\sum_{z=0}^{\infty}\lambda^z + e^{-tar}\sum_{j=r}^{\infty}\lambda^j + \sum_{j=1}^{r-1}\lambda^j e^{ta(r-2j)}\right\}.$$

Summing the two geometric series and taking the logarithm gives $K(t)$ in (D.11). The exact discrete Laplace formula established by Harrison & Manurangsi (2026) gives $\rho(r) = \mathbb{E}_{P_r} L_r$ as displayed. Substitution into the gap definition (3.1) proves the general-t formula.

For part (b), completing the square in the discrete Gaussian sum gives

$$\begin{aligned} e^{K(t)} &= e^{r^2 t(t+1)/(2\sigma^2)}\,\frac{\Theta_\sigma(tr)}{\Theta_\sigma(0)} \\ &= e^{r^2 t(t+1)/(2\sigma^2)}\,\frac{\Theta_\sigma((1+t)r)}{\Theta_\sigma(0)}, \end{aligned}$$

where the second equality uses integer periodicity. Thus

$$K(t) = \frac{r^2}{2\sigma^2}t(t+1) + \log\frac{\Theta_\sigma((1+t)r)}{\Theta_\sigma(0)}.$$

The theta sum is maximized at integers, with equality only at integers. The logarithmic term is therefore nonpositive, while $t = 1/r$ makes $(1+t)r = r+1$ an integer and attains the quadratic envelope. This proves $\rho(r) = r^2/(2\sigma^2)$; differentiating at $t = 0$ also gives $\mathbb{E}_{P_r} L_r = r^2/(2\sigma^2)$. Periodicity of the theta sum gives the stated period and zero set. Poisson summation gives, uniformly in u,

$$\Theta_\sigma(u) = \sqrt{2\pi}\sigma\{1 + 2e^{-2\pi^2\sigma^2}\cos(2\pi u) + O(e^{-8\pi^2\sigma^2})\}.$$

Expanding the logarithm proves (D.12). □