Comparing Poisson and Gaussian channels (extended)

Anzo Teh and Yury Polyanskiy
EECS, MIT
Cambridge, Massachusetts
USA
{anzoteh, yp}@mit.edu

Abstract—Consider a pair of input distributions which after passing through a Poisson channel become ϵ-close in total variation. We show that they must necessarily then be $\epsilon^{0.5+o(1)}$-close after passing through a Gaussian channel as well. In the opposite direction, we show that distributions inducing ϵ-close outputs over the Gaussian channel must induce $\epsilon^{1+o(1)}$-close outputs over the Poisson. This quantifies a well-known intuition that "smoothing" induced by Poissonization and Gaussian convolution are similar. As an application, we improve a recent upper bound of Han-Miao-Shen'2021 for estimating mixing distribution of a Poisson mixture in Gaussian optimal transport distance from $n^{-0.1+o(1)}$ to $n^{-0.25+o(1)}$.

I. INTRODUCTION

Fix three positive parameters $a, \sigma, \gamma > 0$ and consider *two channels* with a common input space $\mathcal{X} = [0, a]$. The first channel, denoted Gsn_σ, acts on input $X = x_0$ by outputting $Y_G \sim \mathcal{N}(x_0, \sigma^2)$. The second channel, denoted Poi_γ, acts by outputting $Y_P \sim (Poi)(\gamma x_0)$. Note that the output spaces of these two channels are very different. For the first one $Y_G \in \mathbb{R}$ and for the second one $Y_P \in \mathbb{Z}_+$. When $X \sim \pi$ we denote by $\mathsf{Gsn}_\sigma \circ \pi$ and $\mathsf{Poi}_\gamma \circ \pi$ the laws of Y_G and Y_P, respectively. Despite the fact that these probability measures live on different spaces, we can view either of them as a kind of "smoothed" version of π, which destroys small local variations in π. One may wonder, thus, whether one can perturb a fixed π in such a way that the perturbation, while invisible after passing through Poisson channel, is apparent after passing through the Gaussian one. In this work, we answer this in the negative and provide quantitive bounds. Specifically, we show that whenever two measures π_1 and π_2 have total variation distance ϵ after Poisson smoothing, they must necessarily also be close after Gaussian smoothing (within total variation almost $O(\sqrt{\epsilon})$), and an even better bound in the opposite direction. Informally speaking, this demonstrates that the information embedded in local variations of X is destroyed similarly by both channels.

Besides independent interest, our results have various applications. One could be in the domain of *covert communication* [1], where coded distribution is supposed to have low total variation distance from a pure noise (our result compares these tasks over two channels). However, our original motivation lies in the domain of *Gaussian optimal transport* (GOT) introduced

in [2]. We recall that a σ-GOT distance is defined as

$$
\begin{aligned}
& W_1^{(\sigma)}(\nu, \mu) \\
& = \inf_{P_{A,B}} \left\{ \mathbb{E}[|A - B|] : A \sim \mathsf{Gsn}_\sigma \circ \nu, B \sim \mathsf{Gsn}_\sigma \circ \mu \right\}, \quad (1)
\end{aligned}
$$

with infimum over all possible joint distributions $P_{A,B}$ with given marginals. When $\sigma = 0$ this corresponds to the standard Wasserstein distance and is denoted by W_1 without the superscript. It is known that estimating a distribution (supported on $[0,1]^d$) in Wasserstein distance is rather slow (typically, at rate $n^{-1/d}$ from n iid samples). If, however, one is interested in only recovering distribution up to features of scale σ, then estimation metric could arguably be replaced by $W_1^{(\sigma)}$. It turns out that estimating in the latter can be done at much faster rates.

One example of this phenomena, and a second motivation for this work, is a result of [3], who showed that estimating π from n iid samples of $\mathsf{Poi}_\gamma \circ \pi$ while essentially impossible [4] in W_1 (rate being $\mathrm{polylog}(n)$) can be done in GOT at a polynomial rate of (almost) $n^{-0.1}$. Our channel comparison analysis paired with a recent bound of [5] improves the estimate to (almost) $n^{-1/4}$.

From the technical side, our innovation is bringing the complex-analytic tools, previously used for Poisson-type problems in [6]–[9] to bear on this channel comparison question. With this brief outline, we proceed to formal statements next.

Notation. $\lesssim$ and $\gtrsim$ denote inequalities up to absolute constants (in particular, these constants do not depend on the problem parameters a, σ). Similarly, $O_{a,\sigma,\gamma}(1)$ and $o_{a,\sigma,\gamma}(1)$ denote quantity that stays bounded or vanishes, but depends on a, σ, γ. $\log$ denotes a base-e logarithm. When doing summation or integral, we will denote $\pi(t)$ as the probability mass (or density) function of distribution π at t.

II. MAIN RESULTS

Throughout the paper, we restrict ourselves to priors of bounded support. That is, we denote $\mathcal{P}([0,a])$ the set of all probability distributions supported on $[0,a]$. In addition to $W_1^{(\sigma)}$, W_1 that were already defined, we also recall definition of TV and Hellinger for two distributions P, Q as follows [10, (7.3), (7.5)].

$$
\mathsf{TV}(P, Q) \triangleq \mathbb{E}_Q \left[\left| \frac{dP}{dQ} - 1 \right| \right] = \frac{1}{2} \int |dP - dQ|, \quad (2)
$$

$$H^2(P,Q) \triangleq \mathbb{E}_Q\left[\left(1 - \sqrt{\frac{dP}{dQ}}\right)^2\right] = \int (\sqrt{dP} - \sqrt{dQ})^2 \quad (3)$$

A. Comparison of Poisson and Gaussian channels

Theorem 1: There exists $c = c(a,\sigma,\gamma) > 0$ such that for any $\pi_1, \pi_2 \in \mathcal{P}([0,a])$ we have

$$\mathsf{TV}(\mathsf{Poi}_\gamma \circ \pi_1, \mathsf{Poi}_\gamma \circ \pi_2) \leq \epsilon$$
$$\implies \mathsf{TV}(\mathsf{Gsn}_\sigma \circ \pi_1, \mathsf{Gsn}_\sigma \circ \pi_2) \leq c(a,\sigma,\gamma)\sqrt{\epsilon}\, t_{a,\sigma,\gamma}(\epsilon)$$

where $t_{a,\sigma}(\epsilon) = \epsilon^{o(1)}$ as $\epsilon \to 0$, and more explicitly, we have

$$t_{a,\sigma,\gamma}(\epsilon) = \frac{\ell_\epsilon^{3/4}}{\sqrt{\log \ell_\epsilon}} e^{\left(\frac{\log a - \log(\sigma^2) - \log(\gamma)}{2} + o(1)\right)\frac{\ell_\epsilon}{\log \ell_\epsilon}}, \ \ell_\epsilon = \log\frac{1}{\epsilon}.$$

In the opposite direction, we have the following result.

Theorem 2: There exists $c = c(a,\sigma,\gamma) > 0$ such that for any $\pi_1, \pi_2 \in \mathcal{P}([0,a])$ we have

$$\mathsf{TV}(\mathsf{Gsn}_\sigma \circ \pi_1, \mathsf{Gsn}_\sigma \circ \pi_2) \leq \epsilon$$
$$\implies \mathsf{TV}(\mathsf{Poi}_\gamma \circ \pi_1, \mathsf{Poi}_\gamma \circ \pi_2) \leq c\epsilon e^{3\gamma\sigma\sqrt{2\log\frac{1}{\epsilon}}}.$$

Remark 1: We consider a simple example on how TVs of Gaussian and Poisson mixtures behave. Let $\pi_1 = \delta_t$ and $\pi_2 = \delta_{t+\epsilon}$ for some small $\epsilon > 0$, and $0 < t < a - \epsilon$, where δ is the dirac-delta distribution. Then $\mathsf{TV}(\mathcal{N}(t,1), \mathcal{N}(t+\epsilon,1)) = \epsilon \cdot (\frac{1}{\sqrt{2\pi}} + o(1))$; while

$$\frac{\exp(-t)}{2}(1 - \exp(-\epsilon)) \overset{(a)}{\leq} \mathsf{TV}(\mathsf{Poi}(t), \mathsf{Poi}(t+\epsilon)) \overset{(b)}{\leq} \epsilon$$

with (a) by comparing the PMF at 0 and (b) by [11, (2.2)]. Since the TV of both channels are of $\Theta(\epsilon)$, the exponent of ϵ in Theorem 1 and Theorem 2 cannot exceed 1.

B. Application to Gaussian optimal transport

Next, we discuss statistical applications of the results above. Consider the problem of estimating the distribution π supported on $[0,a]$ from n iid indirect observations $Y_P \sim \mathsf{Poi}(X)$, $X \sim \pi$. Here we denote the shorthand notation $\mathsf{Poi} \triangleq \mathsf{Poi}_1$. One can pose different questions related to estimating π. For example, while estimating π in TV is impossible, it can be estimated, for example, in Wasserstein W_1 distance, albeit at a slow rate. Specifically, [3] and [4] show that

$$\inf_{\hat{\pi}} \sup_{\pi \in \mathcal{P}([0,a])} \mathbb{E}[W_1(\pi, \hat{\pi})] = \Theta_a\left(\frac{\log\log n}{\log n}\right). \quad (4)$$

Despite the poor performance of estimation of mixing distribution π in this nonparametric inverse problem, estimation of the mixture distribution $\mathsf{Poi} \circ \pi$ can be done at an almost parametric rate. Several different estimators $\hat{\pi}$, including Non-Parametric Maximum Likelihood Estimator (NPMLE), minimum Hellinger distance and minimum χ^2-distance, were shown by [5] to achieve an estimation rate (in Hellinger distance) given by

$$\sup_{\pi \in \mathcal{P}([0,a])} \mathbb{E}[H^2(\mathsf{Poi} \circ \pi, \mathsf{Poi} \circ \hat{\pi})] \leq O_a\left(\frac{\log n}{n\log\log n}\right), \quad (5)$$

where the Hellinger squared distance H^2 was defined above. It was shown previously in [12, Appendix E] that this estimation rate cannot be improved. A previous result by [13, Proposition 3.1] also shows convergence of $\chi(\mathsf{Poi} \circ \hat{\pi} \| \mathsf{Poi} \circ \pi)$ at the rate of faster than $n^{-(1/2-\tau)}$ for any $\tau > 0$.

Finally, estimation under the GOT distance (1) was considered recently. Specifically, [3, Theorem 3.1.] states that for any $0 < c < 0.1$, there exists a constant $C = C(\sigma, a, c)$ such that the NPMLE solution $\hat{\pi}$ satisfies

$$\sup_{\pi \in \mathcal{P}([0,a])} \mathbb{E}[W_1^{(\sigma)}(\pi, \hat{\pi})] \leq Cn^{-c}. \quad (6)$$

We note that [3] presents results for non-Poisson channels as well, but for the Poisson channel $c = 0.1 - o(1)$ is the rate obtained therein, cf. [3, Remark 3.2].

Here we improve this result as follows.

Corollary 2.1: For any $0 < c < \frac{1}{4}$, there exists a constant $C = C(\sigma, a, c)$ such that the NPMLE solution $\hat{\pi}$ of the Poisson mixture attains rate

$$\sup_{\pi \in \mathcal{P}([0,a])} \mathbb{E}[W_1^{(\sigma)}(\pi, \hat{\pi})] \leq Cn^{-c}. \quad (7)$$

Furthermore, if $a \leq \sigma^2\gamma$, then the right-hand side can be replaced with $n^{-1/4}\mathsf{Polylog}(n)$.

We only sketch the main steps here; the complete proof is in Appendix A. First, by the standard bounds, e.g. [10, (7.20)] we have $\mathsf{TV}(P,Q) \leq H(P,Q)$ and thus by Cauchy-Schwarz and (5) we have

$$\sup_{\pi \in \mathcal{P}([0,a])} \mathbb{E}[\mathsf{TV}(\mathsf{Poi} \circ \pi, \mathsf{Poi} \circ \hat{\pi})] = O_a\left(\frac{1}{\sqrt{n}} \cdot \sqrt{\frac{\log n}{\log\log n}}\right)$$

Next, we leverage Theorem 1 to get:

$$\sup_{\pi \in \mathcal{P}([0,a])} \mathbb{E}[\mathsf{TV}(\mathsf{Gsn}_\sigma \circ \hat{\pi}, \mathsf{Gsn}_\sigma \circ \pi)] = O_{a,\sigma}\left(n^{-1/4+o(1)}\right)$$
$$(8)$$

which, in the case where $a \leq \sigma^2\gamma$, $n^{o(1)}$ is actually $\mathsf{Polylog}(n)$ given that $t_{a,\sigma,\gamma}(\epsilon)$ in Theorem 1 is $\mathsf{Polylog}(\frac{1}{\epsilon})$.

For the next step we need the following estimate, to be proven in Appendix A.

Lemma 3: There exists $c_1 = c_1(a,\sigma)$ such that for all $\pi_1, \pi_2 \in \mathcal{P}([0,a])$ and for all $\delta > 0$ we have[1]

$$\mathsf{TV}(\mathsf{Gsn}_\sigma \circ \pi_1, \mathsf{Gsn}_\sigma \circ \pi_2) \leq \delta \implies W_1^{(\sigma)}(\pi_1, \pi_2) \leq c_1\delta\log\frac{1}{\delta}.$$
$$(9)$$

Applying Lemma 3 to (8) we get

$$\sup_{\pi \in \mathcal{P}([0,a])} \mathbb{E}[W_1^{(\sigma)}(\pi, \hat{\pi})] \leq n^{-1/4+o(1)},$$

which completes the proof of Corollary 2.1.

Remark 2 (On the level of smoothing): We have obtained the bound for σ-smoothed distance between NPMLE and truth which is n^{-c} for any $c < 1/4$. This result required a constant fixed σ. However, it turns out that it is sufficient to set $\sigma = 1/\mathsf{Polylog}(n)$, while holding a, γ fixed and

---

[1]We remark that the bound is likely not tight, as for example when $\sigma = 0$, we can easily get a better bound of $W_1(\pi_1, \pi_2) \leq \frac{a\delta}{2}$ [14, Theorem 6.15].

letting $n \to \infty$. Indeed, inspecting the proofs, the constant $c(a, \sigma, \gamma)$ in Theorem 1 is $\exp(\frac{\sigma^2\gamma^2}{2}) \cdot \mathsf{Poly}(\sigma, \frac{1}{\sigma})$. On the other hand, setting $\frac{1}{\sigma^2}$ to grow with $(\log \frac{1}{\epsilon})^v$ where $0 < v < 1$, $t_{a,\sigma,\gamma}(\epsilon)$ becomes $\epsilon^{-v/2}$. Thus, the overall bound in RHS of the Theorem becomes $\epsilon^{(1-v)/2}\mathsf{Polylog}(\epsilon)$. Recalling that $\epsilon = \frac{1}{\sqrt{n}}\sqrt{\frac{\log n}{\log \log n}}$ we get the claimed n^{-c} bound by taking v sufficiently small.

III. COMPLEX-ANALYTIC PRELIMINARIES

The main proof technique for this work is complex analysis. Here, we remind that the z-transform $\mathcal{Z}(\pi)(z)$ (for priors π with discrete support), Laplace transform $\mathcal{L}(\pi)(s)$, and characteristic function $\Psi_\pi(t)$ of a distribution π are defined as follows.

$$\mathcal{Z}(\pi)(z) \triangleq \sum_{n=0}^{\infty} \mathsf{PMF}(\pi)(n)z^n \qquad \forall z \in \mathbb{C} \qquad (10)$$

$$\mathcal{L}(\pi)(s) \triangleq \mathbb{E}_{X \sim \pi}[\exp(sX)] \qquad \forall s \in \mathbb{C} \qquad (11)$$

$$\Psi_\pi(t) \triangleq \mathbb{E}_{X \sim \pi}[\exp(itX)] = \mathcal{L}(\pi)(it) \qquad \forall t \in \mathbb{R} \qquad (12)$$

We now consider the following identities for all bounded priors $\pi \in \mathcal{P}([0, a])$: for Poisson mixtures and Gaussian mixtures we have

$$\mathcal{Z}(\mathsf{Poi}_\gamma \circ \pi)(z) = \mathcal{L}(\pi)(\gamma(z-1)) \quad \forall z \in \mathbb{R}; \qquad (13)$$

$$\mathcal{L}(\mathsf{Gsn}_\sigma \circ \pi)(s) = \exp\left(\frac{s^2\sigma^2}{2}\right)\mathcal{L}(\pi)(s) \quad \forall s \in \mathbb{R}. \qquad (14)$$

In addition, the Plancherel's theorem [15, Theorem 2] implies the following:

$$L_2(\mathsf{Gsn}_\sigma \circ \pi_1, \mathsf{Gsn}_\sigma \circ \pi_2)^2$$
$$\triangleq \int_{-\infty}^{\infty} ((\mathsf{Gsn}_\sigma \circ \pi_1)(t) - (\mathsf{Gsn}_\sigma \circ \pi_2)(t))^2 dt$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-\sigma^2 t^2)|\Psi_{\pi_1}(t) - \Psi_{\pi_2}(t)|^2 dt \qquad (15)$$

using the fact that the Fourier transform of the function $f(t) \triangleq e^{\frac{-\sigma^2 t^2}{2}}\Psi_\pi(t)$ is $\hat{f}(u) = 2\pi(\mathsf{Gsn}_\sigma \circ \pi)(2\pi u)$.

We now describe a main idea that we will be using: the Hadamard's three-circle theorem [16, Theorem 12.1] that states the following. Let $x_0 \in \mathbb{C}$, $r_0 < r_1 \in \mathbb{R}$. Consider a function f that is analytic on the annulus $A_{r_0, r_1} \triangleq \{z : r_0 < |z - x_0| < r_1\}$ and continuous everywhere else. Denote $M_{x_0}(r; f) \triangleq \sup_{|z-x_0| \le r} |f(z)|$. Then

$$\log M_{x_0}(r; f) \text{ is a convex function of } \log r. \qquad (16)$$

Finally, we will also frequently use the following tail bound of the Gaussian distribution [17, Theorem 4.7].

$$\mathbb{P}(N(0, \sigma) > T) \le \sqrt{\frac{2}{\pi}}\frac{\sigma \exp(-T^2/(2\sigma^2))}{T}, \quad \forall T > 0. \qquad (17)$$

We will use (16) to bound the difference in characteristic functions of the Gaussian mixtures. Then the L_2 distance can be bounded via (15) and finally the TV distance via Lemma 4.

IV. PROOF OF THEOREM 1

The following lemma shows that it suffices to bound the L_2 distance in establishing Theorem 1.

Lemma 4: Let $\epsilon, a > 0$ be given, π_1 and $\pi_2 \in \mathcal{P}([0, a])$ be such that

$$L_2(\mathsf{Gsn}_\sigma \circ \pi_1, \mathsf{Gsn}_\sigma \circ \pi_2) \le \epsilon \qquad (18)$$

Then

$$\mathsf{TV}(\mathsf{Gsn}_\sigma \circ \pi_1, \mathsf{Gsn}_\sigma \circ \pi_2) \lesssim \epsilon \cdot \sqrt[4]{\sigma^2 \log \frac{1}{\epsilon} + a} \qquad (19)$$

The complete proof is established in Appendix A. The proof idea is to bound the quantity $\int_{-T}^{T} |\mathsf{Gsn}_\sigma \circ \pi_1 - \mathsf{Gsn}_\sigma \circ \pi_2| dt$ using Cauchy–Schwarz inequality, and bound this quantity outside the said interval using (17) and compactness of support.

Here, we consider the following lemma on transforming bounds on Laplace transform into the characteristic function, relying only on the total variation of the Poisson mixtures and the support bound of the priors.

Lemma 5: Let $\pi_1, \pi_2 \in \mathcal{P}([0, a])$ be such that

$$\sup_{|s+\gamma| \le \gamma} |\mathcal{L}(\pi_1)(s) - \mathcal{L}(\pi_2)(s)| \le 2\epsilon. \qquad (20)$$

Denote $R_\epsilon > 1$ a solution of

$$\log(1/\epsilon) = a(R_\epsilon(\log R_\epsilon - \log \gamma - 1) + \gamma). \qquad (21)$$

Then for all $t \in \mathbb{R}$ we have

$$|\Psi_{\pi_1}(t) - \Psi_{\pi_2}(t)| \le 2\min\left(1, \epsilon \cdot \exp(\frac{a}{2}R_\epsilon \log(1 + \frac{t^2}{\gamma^2}))\right). \qquad (22)$$

Proof of Lemma 5: Denote $f(s) = \frac{\mathcal{L}(\pi_1)(s)}{2} - \frac{\mathcal{L}(\pi_2)(s)}{2}$. For all $r > 0$, we consider $M(r) = \sup_{|s+\gamma| \le r} |f(s)|$ as per Fig. 1. Then we have the following estimates for M:

$$M(\gamma) \le \epsilon, \qquad \forall r > \gamma : M(r) \le \exp(a(r - \gamma)) \qquad (23)$$

where the second one is due to the fact that $\pi_1, \pi_2 \in \mathcal{P}([0, a])$ and $\sup_{|s+\gamma| \le r, x \in [0,a]} |\exp(sx)| = \exp(a(r - \gamma))$. Consider, now, the function $g(u) = \log(M(\gamma e^u))$, then we have $g(0) \le -\log(1/\epsilon)$ and for all $u > 0$, $g(u) \le a\gamma(e^u - 1)$.
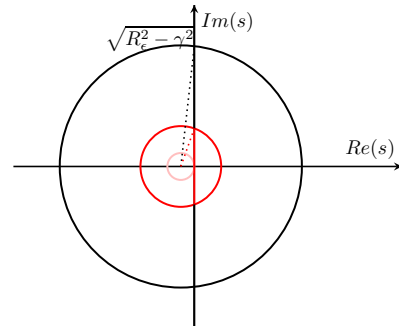

Fig. 1: Bounding $|\Psi_{\pi_1} - \Psi_{\pi_2}|$ on red line using $M(r)$ on pink and black circles.

Given that both π_1 and π_2 are in $\mathcal{P}([0, a])$, f is analytic on $\mathbb{C}$. Therefore, g is convex by (16). Consider R_ϵ as given in
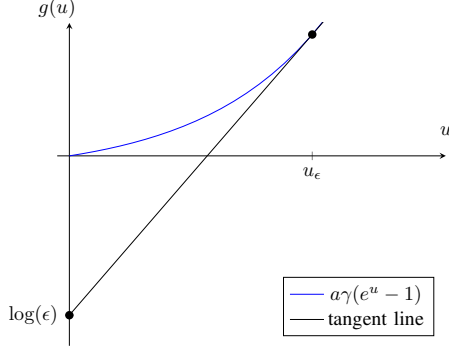
Fig. 2: Bound on $g(u)$ via Hadamard's 3-circle theorem.

(21). Let $u_\epsilon = \log(R_\epsilon) - \log(\gamma)$, then $\log(1/\epsilon) = a\gamma((u_\epsilon - 1)\exp(u_\epsilon) + 1)$. The motivation of this choice of R_ϵ and u_ϵ is given in Fig. 2: for any choice of u_ϵ we would get an upper bound on g given by the line joining the endpoints; the tangent line has the smallest slope, and therefore the best bound.

For each $u \in [0, u_\epsilon]$, the convexity of g entails

$$g(u) \le g(0)(1 - \frac{u}{u_\epsilon}) + g(u_\epsilon) \cdot \frac{u}{u_\epsilon}$$
$$\le -\log(1/\epsilon) + \frac{u}{u_\epsilon}(a\gamma(e^{u_\epsilon} - 1) + a\gamma((u_\epsilon - 1)e^{u_\epsilon} + 1))$$
$$= -\log(1/\epsilon) + a\gamma u \exp(u_\epsilon). \quad (24)$$

Now, $\Psi_{\pi_1}(t) - \Psi_{\pi_2}(t) = 2f(it)$. Since $|\Psi_\pi(t)| \le 1$ for all π, $|f(it)| \le 1$. On the other hand, $|it + \gamma| = \sqrt{\gamma^2 + t^2}$. Therefore, for all $|t| \le \sqrt{R_\epsilon^2 - \gamma^2}$, we have

$$|f(it)| \le M(\sqrt{t^2 + \gamma^2})$$
$$= \exp\left(g(\frac{1}{2}\log(1 + \frac{t^2}{\gamma^2}))\right)$$
$$\le \exp\left(-\log(1/\epsilon) + \frac{a\gamma \exp(u_\epsilon)\log(1 + \frac{t^2}{\gamma^2})}{2}\right)$$
$$= \epsilon \cdot \exp\left(\frac{aR_\epsilon \log(1 + \frac{t^2}{\gamma^2})}{2}\right). \quad (25)$$

On the other hand, for $|t| > \sqrt{R_\epsilon^2 - \gamma^2}$ we have $\epsilon \cdot \exp(\frac{aR_\epsilon \log(1 + \frac{t^2}{\gamma^2})}{2}) \ge \exp(a(R_\epsilon - \gamma)) > 1$, implying that the bound is trivially true then. Thus, $f(it) \le \min(1, \epsilon \cdot \exp(aR_\epsilon \log(1 + t^2)/2))$ for all t. ∎

Proof of Theorem 1: We first establish the following bound via (13).

$$\sup_{s:|s+\gamma|\le\gamma} |\mathcal{L}(\pi_1)(s) - \mathcal{L}(\pi_2)(s)|$$
$$= \sup_{z:|z|\le 1} |\mathcal{Z}(\mathsf{Poi}_\gamma \circ \pi_1)(z) - \mathcal{Z}(\mathsf{Poi}_\gamma \circ \pi_2)(z)|$$
$$\le \sum_{n=0}^{\infty} |(\mathsf{Poi}_\gamma \circ \pi_1)(n) - (\mathsf{Poi}_\gamma \circ \pi_2)(n)|$$
$$= 2\mathsf{TV}(\mathsf{Poi}_\gamma \circ \pi_1, \mathsf{Poi}_\gamma \circ \pi_2) \le 2\epsilon \quad (26)$$

Motivated by (15), we consider R_ϵ as per Lemma 5 and denote $E(s) \triangleq -\sigma^2 s + aR_\epsilon \log(1 + \frac{s}{\gamma^2})$ for all $s > -\gamma^2$. Then

E is concave and attains its global maximum at $s = \frac{aR_\epsilon}{\sigma^2} - \gamma^2$, thus for all $t \in \mathbb{R}$ we have

$$E(t^2) \le E_{\max}$$
$$:= aR_\epsilon(\log(aR_\epsilon) - \log(\sigma^2\gamma^2) - 1) + \sigma^2\gamma^2$$
$$= \log(1/\epsilon) - a\gamma + aR_\epsilon \log(\frac{a}{\sigma^2\gamma}) + \sigma^2\gamma^2. \quad (27)$$

This means we may now bound the squared L_2 distance as follows:

$$L_2(\mathsf{Gsn}_\sigma \circ \pi_1, \mathsf{Gsn}_\sigma \circ \pi_2)^2$$
$$\overset{(a)}{=} \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-\sigma^2 t^2)|\Psi_{\pi_1}(t) - \Psi_{\pi_2}(t)|^2 dt$$
$$\overset{(b)}{\lesssim} \int_{-\infty}^{\infty} \exp(-\sigma^2 t^2) \cdot \min(1, \epsilon \cdot \exp(\frac{aR_\epsilon \log(1 + \frac{t^2}{\gamma^2})}{2}))^2 dt$$
$$\overset{(c)}{\le} \left(\int_{|t|\le R_\epsilon} \epsilon^2 \exp(E(t^2))dt + \int_{|t|>R_\epsilon} \exp(-\sigma^2 t^2)dt\right)$$
$$\overset{(d)}{\lesssim} R_\epsilon \epsilon^2 \exp(E_{\max}) + \exp(-\sigma^2 R_\epsilon^2) \quad (28)$$

where in (a) we used Plancherel (15), in (b) we applied Lemma 5, in (c) we split the integral into two parts and applied respective bounds from previous line, in (d) we used (27) and (17).

To proceed, we notice that the function $f(r) = r\log r$ has $f(\frac{y}{\log y}) = y(1 - \frac{\log\log y}{\log y})$ for all $y > e$, so as $y \to \infty$ the solution to $f(r) = y$ has $r = (1 + o(1))\frac{y}{\log y}$. This, together with (21), implies that $R_\epsilon = (1 + o(1))\frac{1}{a}\frac{\log(1/\epsilon)}{\log\log(1/\epsilon)}$ as $\epsilon \to 0$. Then, the second term in (28) is $o(\epsilon) = o_{a,\sigma,\gamma}(\epsilon)$ and can be neglected, whereas for the first term we can see from (27) that

$$\exp E_{\max} = \frac{1}{\epsilon}s_{a,\sigma,\gamma}(\epsilon),$$

$$s_{a,\sigma,\gamma}(\epsilon) := \exp\left\{\sigma^2\gamma^2 - a\gamma + \left(\log\frac{a}{\sigma^2\gamma} + o(1)\right)\frac{\log\frac{1}{\epsilon}}{\log\log\frac{1}{\epsilon}}\right\}.$$

Collecting terms, thus, we have shown that as $\epsilon \to 0$ we have

$$L_2(\mathsf{Gsn}_\sigma \circ \pi_1, \mathsf{Gsn}_\sigma \circ \pi_2)^2 \lesssim \epsilon R_\epsilon s_{a,\sigma,\gamma}(\epsilon).$$

Finally, taking the square root and invoking Lemma 4 we obtain the statement of the theorem. ∎

V. PROOF OF THEOREM 2

For the comparison in the other direction, we need the following bound on the magnitude of the difference of Laplace transform.

Lemma 6: Consider the same setting as before, where $\pi_1, \pi_2 \in \mathcal{P}([0, a])$. Given $\epsilon > 0$ such that

$$\mathsf{TV}(\mathsf{Gsn}_\sigma \circ \pi_1, \mathsf{Gsn}_\sigma \circ \pi_2) \le \epsilon. \quad (29)$$

Then the Laplace transform satisfies the following:

$$|\mathcal{L}(\pi_1)(s) - \mathcal{L}(\pi_2)(s)|$$
$$\lesssim \epsilon \exp\left(-\frac{\sigma^2 Re(s^2)}{2} + E_{a,\sigma}(\epsilon, s)\right). \quad (30)$$

$$E_{a,\sigma}(\epsilon, s) := \sigma^2 Re(s)^2 + a \cdot |Re(s)| + |Re(s)|\sqrt{2\sigma^2 \log \frac{1}{\epsilon}}.$$
(31)

Proof of Lemma 6: We will show that

$$|\mathcal{L}(\mathsf{Gsn}_\sigma \circ \pi_1 - \mathsf{Gsn}_\sigma \circ \pi_2)(s)| \le \epsilon \exp(E_{a,\sigma}(\epsilon, s)) \quad (32)$$

with $E_{a,\sigma}(\epsilon, s)$ as per (31), and then the conclusion follows from (14).

Indeed, we first consider the following:

$$|\mathcal{L}(\mathsf{Gsn}_\sigma \circ \pi_1 - \mathsf{Gsn}_\sigma \circ \pi_2)(s)|$$
$$\le \int_{-\infty}^{\infty} |\exp(st) \cdot (\mathsf{Gsn}_\sigma \circ \pi_1(t) - \mathsf{Gsn}_\sigma \circ \pi_2(t))| dt$$
$$= \int_{-\infty}^{\infty} \exp(Re(st))|\mathsf{Gsn}_\sigma \circ \pi_1(t) - \mathsf{Gsn}_\sigma \circ \pi_2(t)| dt. (33)$$

Consider $T > \sigma^2|Re(s)| + a$, we now split this into three parts:

$$\int_{-\infty}^{-T} \qquad \int_{-T}^{T} \qquad \int_{T}^{\infty}$$

First, the term in the middle:

$$\int_{-T}^{T} \exp(Re(st))|\mathsf{Gsn}_\sigma \circ \pi_1(t) - \mathsf{Gsn}_\sigma \circ \pi_2(t)| dt$$
$$\le \sup_{|t| \le T} \exp(Re(st)) \int_{-T}^{T} |\mathsf{Gsn}_\sigma \circ \pi_1(t) - \mathsf{Gsn}_\sigma \circ \pi_2(t)| dt$$
$$\le \exp(T \cdot |Re(s)|)\epsilon. \tag{34}$$

Next, for each $\pi \in \mathcal{P}([0, a])$, $\mathsf{Gsn}_\sigma \circ \pi(t)$ is nonnegative for all t, while also bounded above by $\frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{t^2}{2\sigma^2})$ for $t \le 0$, and $\frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(t-a)^2}{2\sigma^2})$ for $t \ge a$. Therefore, denoting:

$$M_1(T) \triangleq T + \sigma^2 Re(s), \qquad M_2(T) \triangleq T - a - \sigma^2 Re(s), \tag{35}$$

the left tail can be computed as

$$\int_{-\infty}^{-T} \exp(Re(st))|\mathsf{Gsn}_\sigma \circ \pi_1(t) - \mathsf{Gsn}_\sigma \circ \pi_2(t)| dt$$
$$\le \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{-T} \exp(tRe(s)) \cdot \exp(-\frac{t^2}{2\sigma^2}) dt$$
$$\overset{(a)}{=} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{-T} \exp(-\frac{(t - \sigma^2 Re(s))^2}{2\sigma^2} + \frac{\sigma^2 Re(s)^2}{2}) dt$$
$$\overset{(b)}{\le} \sqrt{\frac{2}{\pi}} \frac{\sigma}{M_1(T)} \exp(\frac{\sigma^2 Re(s)^2}{2}) \cdot \exp(-\frac{M_1(T)^2}{2\sigma^2})$$

where (a) is completing the square and (b) follows from (17).

We also have the right tail computed similarly as

$$\int_{T}^{\infty} \exp(Re(st))|\mathsf{Gsn}_\sigma \circ \pi_1(t) - \mathsf{Gsn}_\sigma \circ \pi_2(t)| dt$$
$$\le \frac{1}{\sqrt{2\pi}\sigma} \int_{T}^{\infty} \exp(tRe(s)) \cdot \exp(-\frac{(t-a)^2}{2\sigma^2}) dt$$
$$\le \sqrt{\frac{2}{\pi}} \frac{\sigma}{M_2(T)} \exp(\frac{\sigma^2 Re(s)^2}{2} + a \cdot Re(s)) \cdot \exp(-\frac{M_2(T)^2}{2\sigma^2}).$$

Denote, now, $M_3(T) \triangleq T - a - \sigma^2|Re(s)|$. Then $\min\{M_1(T), M_2(T)\} \ge M_3(T) \ge 0$. Therefore, collecting terms above,

$$|\mathcal{L}(\mathsf{Gsn}_\sigma \circ \pi_1 - \mathsf{Gsn}_\sigma \circ \pi_2)(s)|$$
$$\lesssim \epsilon \exp(T \cdot |Re(s)|) \tag{36}$$
$$+ \frac{\sigma}{M_3(T)} \exp(-\frac{M_3(T)^2}{2\sigma^2} + \frac{\sigma^2 Re(s)^2}{2} + a \cdot |Re(s)|). (37)$$

Next, we choose $T = \sigma^2|Re(s)| + a + \sqrt{2\sigma^2 \log \frac{1}{\epsilon}}$. Thus, the first term (36) evaluates to

$$\epsilon \exp(\sigma^2 Re(s)^2 + a \cdot |Re(s)| + |Re(s)|\sqrt{2\sigma^2 \log \frac{1}{\epsilon}}).$$

With this choice of T, we have $M_3(T) = \sqrt{2\sigma^2 \log \frac{1}{\epsilon}}$. Then, the second term (37) is bounded as

$$\frac{\sigma}{\sqrt{2\sigma^2 \log(\frac{1}{\epsilon})}} \exp(\frac{\sigma^2 Re(s)^2}{2} + a \cdot |Re(s)|)\epsilon.$$

Therefore collecting the two terms together, and taking the maximum of the exponents, gives us (32). ∎

Proof of Theorem 2: As in [7, (33)] we use the standard fact that for any real $r > 1$, a function $f(z) \triangleq \sum_{n=0}^{\infty} a_n z^n$ satisfies

$$\sum_{n=0}^{\infty} |a_n| \le \frac{r}{r-1} \sup_{|z| \le r} |f(z)| \tag{38}$$

(as a consequence of Cauchy's integral formula).

Now if $a_n = (\mathsf{Poi} \circ \pi_1)(n) - (\mathsf{Poi} \circ \pi_2)(n)$, then using (13), $f(z) = \mathcal{L}(\pi_1)(\gamma(z-1)) - \mathcal{L}(\pi_2)(\gamma(z-1))$. Thus setting $r = 2$, we have, by Lemma 6,

$$2\mathsf{TV}(\mathsf{Poi} \circ \pi_1, \mathsf{Poi} \circ \pi_2)$$
$$\le 2 \sup_{|z| \le r} |f(z)|$$
$$= 2 \sup_{|z| \le r} |\mathcal{L}(\pi_1)(\gamma(z-1)) - \mathcal{L}(\pi_2)(\gamma(z-1))|$$
$$\lesssim \sup_{|s+\gamma| \le 2\gamma} \epsilon \exp\left(-\frac{\sigma^2 Re(s^2)}{2} + E_{a,\sigma}(\epsilon, s)\right)$$
$$\le \epsilon \exp\left(3\gamma\sigma \cdot \sqrt{2\log \frac{1}{\epsilon}} + \frac{9\gamma^2\sigma^2}{2} + 9\gamma^2\sigma^2 + 3\gamma a\right)$$

where we used $|Re(s)| \le |s| \le 3\gamma$ and $|Re(s^2)| \le |s^2| \le 9\gamma^2$ for all s with $|s+\gamma| \le 2\gamma$. ∎

REFERENCES

[1] B. A. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on awgn channels," *IEEE journal on selected areas in communications*, vol. 31, no. 9, pp. 1921–1930, 2013.

[2] Z. Goldfeld, K. Greenewald, Y. Polyanskiy, and J. Weed, "Convergence of Smoothed Empirical Measures with Applications to Entropy Estimation," in *IEEE Trans. Inf. Theory*, vol. 66, no. 7, Jul. 2020, pp. 4368–4391.

[3] F. Han, Z. Miao, and Y. Shen, "Nonparametric mixture MLEs under Gaussian-smoothed optimal transport distance," *arXiv preprint arXiv:2112.02421*, Dec. 2021.

[4] Z. Miao, W. Kong, R. K. Vinayak, W. Sun, and F. Han, "Fisher-Pitman permutation tests based on nonparametric Poisson mixtures with application to single cell genomics," *arXiv preprint arXiv:2106.03022*, Jun. 2021.

[5] S. Jana, Y. Polyanskiy, and Y. Wu, "Optimal empirical Bayes estimation for the Poisson model via minimum-distance methods," *arXiv preprint arXiv:2209.01328*, Sep. 2022.

[6] A. Moitra and M. Saks, "A polynomial time algorithm for lossy population recovery," in *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*. IEEE, 2013, pp. 110–116.

[7] Y. Polyanskiy, A. T. Suresh, and Y. Wu, "Sample complexity of population recovery," in *Proceedings of the 2017 Conference on Learning Theory*, ser. Proceedings of Machine Learning Research, S. Kale and O. Shamir, Eds., vol. 65. PMLR, 07–10 Jul 2017, pp. 1589–1618. [Online]. Available: https://proceedings.mlr.press/v65/polyanskiy17a.html

[8] Y. Polyanskiy and Y. Wu, "Dualizing Le Cam's method for functional estimation, with applications to estimating the unseens," *arXiv preprint arXiv:1902.05616*, Feb. 2019.

[9] S. Jana, Y. Polyanskiy, and Y. Wu, "Extrapolating the profile of a finite population," in *Conference on Learning Theory*. PMLR, 2020, pp. 2011–2033.

[10] Y. Polyanskiy and Y. Wu, *Information Theory: From Coding to Learning*. Cambridge University Press, 2022+. [Online]. Available: https://people.lids.mit.edu/yp/homepage/data/itbook-export.pdf

[11] J. A. Adell and P. Jodrá, "Exact Kolmogorov and total variation distances between some familiar discrete distributions," *Journal of Inequalities and Applications*, vol. 2006, no. 1, pp. 1–8, Dec. 2006, number: 1 Publisher: SpringerOpen.

[12] Y. Polyanskiy and Y. Wu, "Sharp regret bounds for empirical Bayes and compound decision problems," *arXiv preprint arXiv:2109.03943*, Sep. 2021.

[13] D. Lambert and L. Tierney, "Asymptotic Properties of Maximum Likelihood Estimates in the Mixed Poisson Model," *The Annals of Statistics*, vol. 12, no. 4, pp. 1388–1399, Dec. 1984, publisher: Institute of Mathematical Statistics.

[14] C. Villani, *Optimal Transport: Old and New*. Berlin: Springer Verlag, 2008.

[15] N. Wiener, *The Fourier integral and certain of its applications*, ser. Cambridge mathematical library. Cambridge: Cambridge University Press, 1988.

[16] B. Simon, *Convexity: An Analytic Viewpoint*, ser. Cambridge Tracts in Mathematics. Cambridge University Press, 2011.

[17] L. Wasserman, *Inequalities*. New York, NY: Springer New York, 2004, pp. 63–69. [Online]. Available: https://doi.org/10.1007/978-0-387-21736-9_4

APPENDIX A
PROOFS OF AUXILLARY LEMMAS

Proof of Corollary 2.1: We first consider the following steps: there is a constant $c_1 = c_1(a)$ such that

$$\sup_{\pi \in \mathcal{P}([0,a])} \mathbb{E}[\mathsf{TV}(\mathsf{Poi} \circ \pi, \mathsf{Poi} \circ \hat{\pi})]$$

$$\overset{(a)}{\leq} \sup_{\pi \in \mathcal{P}([0,a])} \mathbb{E}[H(\mathsf{Poi} \circ \pi, \mathsf{Poi} \circ \hat{\pi})]$$

$$\overset{(b)}{\leq} \sup_{\pi \in \mathcal{P}([0,a])} \sqrt{\mathbb{E}[H^2(\mathsf{Poi} \circ \pi, \mathsf{Poi} \circ \hat{\pi})]}$$

$$\overset{(c)}{\lesssim} \sup_{\pi \in \mathcal{P}([0,a])} c_1(a) \frac{1}{\sqrt{n}} \cdot \sqrt{\frac{\log n}{\log \log n}} \tag{39}$$

where (a) is due to $\mathsf{TV}(P,Q) \leq H(P,Q)$ [10, (7.20)], (b) is $(\mathbb{E}[H(P,Q)])^2 \leq \mathbb{E}[H^2(P,Q)]$ by Cauchy-Schawrz inequality, and (c) is by (5).

Combining Theorem 1 and Lemma 3, we see that there is a constant $c_2 = c_2(a,\sigma)$ such that for all $\pi, \hat{\pi} \in \mathcal{P}([0,a])$ and $X > 0$,

$$\mathsf{TV}(\mathsf{Poi} \circ \pi, \mathsf{Poi} \circ \hat{\pi}) \leq X \implies W_1^{(\sigma)}(\pi, \hat{\pi}) \leq c_2 \sqrt{X} u_{a,\sigma}(X) \tag{40}$$

where for all x with $0 < x < \frac{1}{2e}$ we define $u_{a,\sigma}(x) \triangleq t_{a,\sigma}(x) \log(\frac{1}{x})$, $t_{a,\sigma}(x)$ as per Theorem 1.

Now we have two cases:

- If $a < \sigma^2$, then $\lim u_{a,\sigma}(x) \to 0$ as $x \to 0$, (the polylog factor of $\frac{1}{x}$ is offset by the factor $\exp((\frac{\log a - \log \sigma^2}{2} + o(1))\frac{\log \frac{1}{x}}{\log \log \frac{1}{x}}))$ so $u_{a,\sigma}(x)$ is bounded in $(0, \frac{2}{e})$ by some factor $C = C(a,\sigma)$.

- If $a \geq \sigma^2$, then there is $N = N(a,\sigma^2)$ such that $u_{a,\sigma}(\frac{1}{\sqrt{n}})$ is increasing in n but $\frac{1}{\sqrt[4]{n}} u_{a,\sigma}(\frac{1}{\sqrt{n}})$ is decreasing in n for $n \geq N$. This means, when $X \geq \frac{1}{\sqrt{n}}$, $u_{a,\sigma}(X) \leq u_a(\frac{1}{\sqrt{n}})$; when $X \leq \frac{1}{\sqrt{n}}$, $\sqrt{X} u_{a,\sigma}(X) \leq \frac{1}{\sqrt[4]{n}} u_a(\frac{1}{\sqrt{n}})$.

The first case gives us

$$\sup_{\pi \in \mathcal{P}([0,a])} \mathbb{E}[W_1^{(\sigma)}(\pi, \hat{\pi})] \leq C \mathbb{E}[\sqrt{X}]$$

$$\overset{(a)}{\leq} C \sqrt{\mathbb{E}[X]} \overset{(b)}{\lesssim} C c_2 (\frac{\log n}{n \log \log n})^{1/4}$$

where (a) follows from Cauchy-Schwarz inequality and (b) from (39).

For the second case, we have

$$\sup_{\pi \in \mathcal{P}([0,a])} \mathbb{E}[W_1^{(\sigma)}(\pi, \hat{\pi})] \overset{(a)}{\leq} \mathbb{E}[\sqrt{X} u_{a,\sigma}(X)]$$

$$= \mathbb{E}[\sqrt{X} u_{a,\sigma}(X) \mathbf{1}_{\left\{X \geq \frac{1}{\sqrt{n}}\right\}}]$$

$$+ \mathbb{E}[\sqrt{X} u_{a,\sigma}(X) \mathbf{1}_{\left\{X < \frac{1}{\sqrt{n}}\right\}}]$$

$$\leq u_{a,\sigma}\left(\frac{1}{\sqrt{n}}\right) \mathbb{E}[\sqrt{X}] + \frac{1}{\sqrt[4]{n}} u_{a,\sigma}\left(\frac{1}{\sqrt{n}}\right)$$

$$\overset{(b)}{\lesssim} c_2 \left(\frac{\log n}{n \log \log n}\right)^{1/4} u_{a,\sigma}\left(\frac{1}{\sqrt{n}}\right) \tag{41}$$

where the (a) follows from (40), and (b) from Cauchy-Schawrz and (39). Finally,

$$u_{a,\sigma}(\frac{1}{\sqrt{n}})$$

$$= \frac{\log(\sqrt{n})^{7/4}}{\sqrt{\log \log (\sqrt{n})}} \exp((\frac{\log a - \log(\sigma^2)}{2} + o(1))\frac{\log(\sqrt{n})}{\log \log (\sqrt{n})})$$

which is $n^{o_{a,\sigma}(1)}$ as $n \to \infty$. Therefore $\mathbb{E}[W_1^{(\sigma)}(\pi, \hat{\pi})] \lesssim n^{-1/4 + o_{a,\sigma}(1)}$. ∎

Proof of Lemma 3: We consider the following statement in [14, Theorem 6.15] (using $p = 1, p' = \infty$) : For any point x_0 we have

$$W_1(\pi_1, \pi_2) \leq \int_{x=-\infty}^{\infty} |x_0 - x||d\pi_1(x) - d\pi_2(x)|. \quad (42)$$

Choose $x_0 = \frac{a}{2}$, then for every $T > a$ we have

$$
\begin{aligned}
&W_1^{(\sigma)}(\pi_1, \pi_2) \\
&\leq \int_{-\infty}^{\infty} |u - \frac{a}{2}| \cdot |(\mathsf{Gsn}_\sigma \circ \pi_1)(u) - (\mathsf{Gsn}_\sigma \circ \pi_2)(u)| du \\
&\leq \int_{|u-\frac{a}{2}| \leq T-\frac{a}{2}} |u - \frac{a}{2}| \cdot |(\mathsf{Gsn}_\sigma \circ \pi_1)(u) - (\mathsf{Gsn}_\sigma \circ \pi_2)(u)| du \\
&\quad + \int_{|u-\frac{a}{2}| > T-\frac{a}{2}} |u - \frac{a}{2}| \cdot |(\mathsf{Gsn}_\sigma \circ \pi_1)(u) - (\mathsf{Gsn}_\sigma \circ \pi_2)(u)| du \\
&\leq 2\delta |T - \frac{a}{2}|^2 \\
&\quad + \int_{|u-\frac{a}{2}| > T-\frac{a}{2}} |u - \frac{a}{2}| \cdot |(\mathsf{Gsn}_\sigma \circ \pi_1)(u) - (\mathsf{Gsn}_\sigma \circ \pi_2)(u)| du.
\end{aligned}
$$
$$(43)$$

Because π_1 and π_2 are supported on $[0, a]$, for $\pi \in \{\pi_1, \pi_2\}$ and for all u with $|u - \frac{a}{2}| > \frac{a}{2}$ we have

$$0 \leq \mathsf{Gsn}_\sigma \circ \pi(u) \leq \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(|u - a/2| - a/2)^2}{2\sigma^2}).$$

Now that the tail bound is symmetric on both sides, we have

$$
\begin{aligned}
&\int_{|u-\frac{a}{2}| > T-\frac{a}{2}} |u - \frac{a}{2}| \cdot |(\mathsf{Gsn}_\sigma \circ \pi_1)(u) - (\mathsf{Gsn}_\sigma \circ \pi_2)(u)| du \\
&\leq \frac{2}{\sqrt{2\pi}\sigma} \int_T^{\infty} |u - \frac{a}{2}| \exp(-\frac{(|u - a/2| - a/2)^2}{2\sigma^2}) du \\
&= \frac{2}{\sqrt{2\pi}\sigma} \int_T^{\infty} (u - \frac{a}{2}) \exp(-\frac{(u - a)^2}{2\sigma^2}) du \\
&\overset{(a)}{=} a\mathbb{P}[N(0, \sigma^2) > T - a] + \frac{2\sigma}{\sqrt{2\pi}} \exp(-\frac{(T - a)^2}{\sqrt{2\pi}}) \\
&\overset{(b)}{\leq} \left(\frac{\sqrt{2}a\sigma}{\sqrt{\pi}(T - a)} + \frac{2\sigma}{\sqrt{2\pi}} \right) \exp(-\frac{(T - a)^2}{2\sigma^2})
\end{aligned}
$$
$$(44)$$

where (a) is by the expansion of $(u - \frac{a}{2}) \exp(-\frac{(u-a)^2}{2\sigma^2})$ into $\frac{a}{2} \exp(-\frac{(u-a)^2}{2\sigma^2}) + (u - a) \exp(-\frac{(u-a)^2}{2\sigma^2})$, and (b) (first term) is due to [17, Theorem 4.7].

Finally, setting $T = \sqrt{2\sigma^2 \log(1/\delta)} + a$, (43) is now bounded by

$$
\begin{aligned}
&2\left(\sqrt{2\sigma^2 \log(1/\delta)} + \frac{a}{2} \right)^2 \delta \\
&+ \left(\frac{\sqrt{2}a\sigma}{\sqrt{\pi}\sqrt{2\sigma^2 \log(1/\delta)}} + \frac{2\sigma}{\sqrt{2\pi}} \right) \delta \\
&\lesssim \delta \left(2\sigma^2 \log(1/\delta) + a^2 + a + \sigma \right).
\end{aligned}
$$

∎

Proof of Lemma 4: According to the definition of L2, we have

$$\int_{-\infty}^{\infty} ((\mathsf{Gsn}_\sigma \circ \pi_1)(t) - (\mathsf{Gsn}_\sigma \circ \pi_2(t))^2 dt \leq \epsilon^2. \quad (45)$$

Consider any $T > a$. By Cauchy-Schwarz inequality we have

$$
\begin{aligned}
&\left(\int_{-T}^{T} |(\mathsf{Gsn}_\sigma \circ \pi_1)(t) - (\mathsf{Gsn}_\sigma \circ \pi_2)(t)| dt \right)^2 \\
&\leq \left(\int_{-T}^{T} ((\mathsf{Gsn}_\sigma \circ \pi_1)(t) - (\mathsf{Gsn}_\sigma \circ \pi_2)(t))^2 dt \right) \left(\int_{-T}^{T} 1 dt \right) \\
&\lesssim 2T \cdot \epsilon^2.
\end{aligned}
$$
$$(46)$$

In addition, since both π_1 and $\pi_2 \in \mathcal{P}([0, a])$, we have $\mathbb{P}(|\mathsf{Gsn}_\sigma \circ \pi| > T) \leq 2\mathbb{P}(N(0, \sigma) > T - a) \lesssim \frac{2\sigma}{T-a} \exp(-\frac{(T-a)^2}{2\sigma^2})$ for each $\pi \in \{\pi_1, \pi_2\}$, with the last inequality follows from (17). This means

$$
\begin{aligned}
&\int_{|t|>T} |(\mathsf{Gsn}_\sigma \circ \pi_1)(t) - (\mathsf{Gsn}_\sigma \circ \pi_2)(t)| dt \\
&\leq \int_{|t|>T} |(\mathsf{Gsn}_\sigma \circ \pi_1)(t)| + |(\mathsf{Gsn}_\sigma \circ \pi_2)(t)| dt \\
&\lesssim \frac{2\sigma}{T - a} \exp(-\frac{(T - a)^2}{2\sigma^2}).
\end{aligned}
$$
$$(47)$$

Thus collecting (46) and (47) we have

$$
\begin{aligned}
&2\mathsf{TV}((\mathsf{Gsn}_\sigma \circ \pi_1)(t), (\mathsf{Gsn}_\sigma \circ \pi_2)(t)) \\
&\lesssim \sqrt{T} \cdot \epsilon + \frac{\sigma}{T - a} \exp(-\frac{(T - a)^2}{2\sigma^2}).
\end{aligned}
$$

Now choose $T = \sqrt{2\sigma^2 \log(\frac{1}{\epsilon})} + a$, we have

$$
\begin{aligned}
&\sqrt{T} \cdot \epsilon + \frac{1}{T - a} \exp(-\frac{(T - a)^2}{2}) \\
&= \epsilon \sqrt{\sqrt{2\sigma^2 \log(\frac{1}{\epsilon})} + a} + \frac{\epsilon}{\sqrt{2 \log(\frac{1}{\epsilon})}} \\
&\lesssim \epsilon \cdot \sqrt[4]{\sigma^2 \log(\frac{1}{\epsilon}) + a}
\end{aligned}
$$

where the last inequality we used $\sqrt{2\sigma^2 \log(\frac{1}{\epsilon})} + a \leq \sqrt{2(2\sigma^2 \log(\frac{1}{\epsilon}) + a^2)}$.

∎