Rate of convergence of the smoothed empirical Wasserstein distance

Adam Block, Zeyu Jia, Yury Polyanskiy, Alexander Rakhlin
Massachusetts Institute of Technology

May 4, 2022

Abstract

Consider an empirical measure $\mathbb{P}_n$ induced by n iid samples from a d-dimensional K-subgaussian distribution $\mathbb{P}$ and let $\gamma = \mathcal{N}(0, \sigma^2 I_d)$ be the isotropic Gaussian measure. We study the speed of convergence of the smoothed Wasserstein distance $W_2(\mathbb{P}_n * \gamma, \mathbb{P} * \gamma) = n^{-\alpha+o(1)}$ with $*$ being the convolution of measures. For $K < \sigma$ and in any dimension $d \geq 1$ we show that $\alpha = \frac{1}{2}$. For $K > \sigma$ in dimension $d = 1$ we show that the rate is slower and is given by $\alpha = \frac{(\sigma^2+K^2)^2}{4(\sigma^4+K^4)} < 1/2$. This resolves several open problems in [GGNWP20], and in particular precisely identifies the amount of smoothing σ needed to obtain a parametric rate. In addition, we also establish that $D_{KL}(\mathbb{P}_n * \gamma \| \mathbb{P} * \gamma)$ has rate $O(1/n)$ for $K < \sigma$ but only slows down to $O(\frac{(\log n)^{d+1}}{n})$ for $K > \sigma$. The surprising difference of the behavior of W_2^2 and KL implies the failure of T_2-transportation inequality when $\sigma < K$. Consequently, the requirement $K < \sigma$ is necessary for validity of the log-Sobolev inequality (LSI) for the Gaussian mixture $\mathbb{P} * \mathcal{N}(0, \sigma^2)$, closing an open problem in [WW$^+$16], who established the LSI under precisely this condition.

1 Introduction and main results

Given n iid samples $X_1, \ldots, X_n$ from a probability measure $\mathbb{P}$ on $\mathbb{R}^d$ let us denote by $\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ the empirical distribution. As $n \to \infty$ it is well known that $\mathbb{P}_n \to \mathbb{P}$ according to many different notions of convergence. The literature on the topic is voluminous. Here we are interested in convergence in Wasserstein W_p-distances, cf. [Vil03, Chapter 1], defined for $p \geq 1$ as

$$W_p(\mathbb{P}, \mathbb{Q})^p = \inf_{P_{X,Y}} \left\{ \mathbb{E}[\|X - Y\|^p] : P_X = \mathbb{P}, P_Y = \mathbb{Q} \right\},$$

where $\|\cdot\|$ is Euclidean norm. Already in [Dud69] it was shown that

$$W_1(\mathbb{P}_n, \mathbb{P}) = \Theta(n^{-1/d}),$$

for $d \geq 2$ and compactly supported $\mathbb{P}$ absolutely continuous with respect to Lebesgue measure. Dudley's technique relied on the characterization (special to $p = 1$) of W_1 as the supremum over expectations of Lipschitz functions. His idea of recursive partitioning was cleverly adapted to the realm of couplings in [BLG14], recovering Dudley's convergence rate of $n^{-1/d}$ also for $p > 1$. See [DSS13, FG15, WB19] for more on this line of work, and also for a thorough survey of the recent literature.

Somewhat surprisingly, it was discovered in [GGNWP20] that the rate of convergence improves all the way to (dimension-independent) $n^{-1/2}$ if one merely regularizes both $\mathbb{P}_n$ and $\mathbb{P}$ by convolving with the Gaussian density.[1] More precisely, let $\varphi_{\sigma^2}(x) \triangleq (2\pi\sigma^2)^{-d/2} e^{-\frac{\|x\|^2}{2\sigma^2}}$ be the density of $\mathcal{N}(0, \sigma^2 I_d)$, and for any probability measure $\mathbb{P}$ on $\mathbb{R}^d$ we define the convolved measure via

$$\mathbb{P} * \mathcal{N}(0, \sigma^2 I_d)(E) = \int_E dz \mathbb{E}\left[\varphi_{\sigma^2}(X - z)\right], \quad X \sim \mathbb{P},$$

[1] Of course, the price to pay for this fast rate is a constant in front of $n^{-1/2}$, which can be exponential in d for certain $\mathbb{P}$, cf [GGNWP20].

where E is any Borel set. Then [GGNWP20, Prop. 6] shows

$$\mathbb{E}[W_2^2(\mathbb{P}_n * \mathcal{N}(0, \sigma^2 I_d), \mathbb{P} * \mathcal{N}(0, \sigma^2 I_d))] \le \frac{C(d, \sigma, K)}{n}, \tag{1}$$

whenever $\mathbb{P}$ is K-subgaussian and $K < \frac{\sigma}{2}$. We recall that $X \sim \mathbb{P}$ is K-subgaussian if

$$\mathbb{E}[e^{(\lambda, X - \mathbb{E}[X])}] \le e^{\frac{1}{2}K^2\|\lambda\|^2} \qquad \forall \lambda \in \mathbb{R}^d. \tag{2}$$

Note that in (1) constant C does not depend on $\mathbb{P}$. Estimate (1) is most exciting for large d, but even for $d = 1$ and $\mathbb{P} = \mathcal{N}(0, 1)$ it is non-trivial as $\mathbb{E}[W_2^2(\mathbb{P}_n, \mathbb{P})] \asymp \frac{\log \log n}{n}$. Another surprising feature is [GGNWP20, Corollary 2]: for $K \ge \sqrt{2}\sigma$ there exists a K-subgaussian distribution $\mathbb{P}$ in $\mathbb{R}^1$ such that

$$\lim_{n \to \infty} n\mathbb{E}[W_2^2(\mathbb{P}_n * \mathcal{N}(0, \sigma^2 I_d), \mathbb{P} * \mathcal{N}(0, \sigma^2 I_d))] = \infty, \tag{3}$$

where the expectation is with respect to n samples according to $\mathbb{P}$. We say that the rate of convergence is "parametric" if (1) holds and otherwise call it "non-parametric". Thus, the results of [GGNWP20] show that parametric rate for smoothed-W_2 is only attained by sufficiently light-tailed distributions $\mathbb{P}$, e.g. the subgaussian constant of distribution $\mathbb{P}$ is less than the scale of noise over two.

In this paper we prove three principal results:

1. Theorem 1 resolves the gap between the location of the parametric and non-parametric region: it turns out that for $K < \sigma$ we always have (1), while for $K > \sigma$ we have (3) for some K-subgaussian distribution $\mathbb{P}$ in $\mathbb{R}^1$. (We remark that for W_1 we always have parametric rate $n^{-1/2}$ for all $K, \sigma > 0$, cf [GGNWP20, Proposition 1].)

2. In the region of non-parametric rates ($K > \sigma$) a natural question arises: what rates of convergence are possible? In other words, what is the value of

$$\alpha = \alpha(K, \sigma, d) \triangleq \lim_{n \to \infty} \inf_{\mathbb{P} - K\text{-subgaussian}} -\frac{\log \mathbb{E}[W_2(\mathbb{P}_n * \mathcal{N}(0, \sigma^2 I_d), \mathbb{P} * \mathcal{N}(0, \sigma^2 I_d))]}{\log n} \tag{4}$$

   Previously, it was only known that $\frac{1}{4} \le \alpha \le \frac{1}{2}$ for all $K > \sigma$ (note that (3) strongly suggests but does not formally imply $\alpha < \frac{1}{2}$). Theorem 2 shows that for $d = 1$ we have

$$\alpha(K, \sigma, d = 1) = \frac{(\sigma^2 + K^2)^2}{4(\sigma^4 + K^4)}.$$

3. We can see that for a class of K-subgaussian distributions, the convergence rate of $W_2(\mathbb{P}_n * \mathcal{N}(0, \sigma^2 I_d), \mathbb{P} * \mathcal{N}(0, \sigma^2 I_d))$ changes from $n^{-1/4}$ to $n^{-1/2}$ as σ increases from 0 to K, after which the rate remains $n^{-1/2}$. Our final result (Theorem 3) shows that, despite being intimately related to W_2, the Kullback-Leibler (KL) divergence behaves rather differently: For all K-subgaussian $\mathbb{P}$ we have

$$\mathbb{E}[D_{KL}(\mathbb{P}_n * \mathcal{N}(0, \sigma^2 I_d) \| \mathbb{P} * \mathcal{N}(0, \sigma^2 I_d))] \le \begin{cases} \frac{C(\sigma, K, d) \log^d n}{n}, & K > \sigma \\ \frac{C(\sigma, K, d)}{n}, & K < \sigma \end{cases}, \tag{5}$$

   where $D_{KL}(\mu \| \nu) = \int d\nu f(x) \log f(x)$, $f \triangleq \frac{d\mu}{d\nu}$ whenever μ is absolutely continuous with respect to ν. Now from the proof of Theorem 1 we also know that for $K > \sigma$, KL-divergence is $\omega(\frac{1}{n})$. Thus, while at $K > \sigma$ both W_2 and KL switch to the non-parametric regime, the W_2 distance experiences a polynomial slow-down in rate, while KL only gets hit by (at most) a poly-logarithmic penalty.

To better understand the relationship between the W_2 results and the KL one, let us recall an important result of Talagrand (known as T_2-transportation inequality). A probability measure ν is said to satisfy the T_2 inequality if there exists a finite constant C such that

$$\forall \mathbb{Q}: \quad W_2^2(\mathbb{Q}, \nu) \le C \cdot D_{KL}(\mathbb{Q} \| \nu).$$

The infimum over all such constants is denoted by $T_2(\nu)$. Talagrand originally demonstrated that $T_2(\mathcal{N}(0, \sigma^2 I_d)) < \infty$ It turns out that $T_2(\mathbb{P} * \mathcal{N}(0, \sigma^2 I_d)) < \infty$ as well for compactly supported $\mathbb{P}$ [Zim13] and K-subgaussian $\mathbb{P}$ with $K < \sigma$ [WW+16] (in fact, both papers establish a stronger log-Sobolev inequality (LSI)). Also in [CCNW21], sharper LSI constants for distribution $\mathbb{P} * \mathcal{N}(0, \sigma^2 I_d)$ are given where $\mathbb{P}$ is with compact support or subgaussianity.

Now comparing (5) and the lower bound for all $K > \sigma$ established in Theorem 2 we discover the following.

Corollary 1. *For any $K > \sigma$ there exists a K-subgaussian $\mathbb{P}$ on $\mathbb{R}^1$ such that $\mathbb{P} * \mathcal{N}(0, \sigma^2)$ does not satisfy T_2-transportation inequality (and hence does not satisfy the LSI either), that is $T_2(\mathbb{P} * \mathcal{N}(0, \sigma^2)) = \infty$.*

We remark that it is straightforward to show that

$$\sup\{T_2(\mathbb{P} * \mathcal{N}(0, \sigma^2)) : \mathbb{P} - K\text{-subgaussian}\} = \infty$$

by simply considering $\mathbb{P} = (1 - \epsilon)\delta_0 + \epsilon\delta_N$ for $\epsilon \to 0$ and $N \to \infty$ (cf. Appendix B). However, each of these measures has $T_2 < \infty$. Evidently, our corollary proves a stronger claim.

Incidentally, this strengthening resolves an open question stated in [WW+16], who proved the LSI (and T_2) for $\mathbb{P} * \mathcal{N}(0, \sigma^2)$ assuming $\mathbb{E}[e^{aX^2}] < \infty$ holds for some $a > \frac{1}{2\sigma^2}$, where $X \sim \mathbb{P}$. They raised a question whether this threshold can be reduced, and our Corollary shows the answer is negative. Indeed, one only needs to notice that whenever $X \sim \mathbb{P}$ is K-subgaussian it satisfies

$$\mathbb{E}\left[e^{aX^2}\right] < \infty \quad \forall a < \frac{1}{2K^2}, \tag{6}$$

which is proved in [BLM13, p. 26].

1.1 Main results and proof ideas

Our first result is the following:

Theorem 1. *If $K < \sigma$, then for any K-subgaussian distribution $\mathbb{P}$, we have*

$$\mathbb{E}\left[W_2^2(\mathbb{P}_n * \mathcal{N}(0, \sigma^2 I_d), \mathbb{P} * \mathcal{N}(0, \sigma^2 I_d))\right] = \mathcal{O}\left(\frac{1}{n}\right),$$

where $\mathbb{P}_n$ is the empirical measure of $\mathbb{P}$ with n samples, and the expectation is over these n samples. If $K > \sigma$, then there exists a K-subgaussian distribution $\mathbb{P}$ such that

$$\mathbb{E}\left[W_2^2(\mathbb{P}_n * \mathcal{N}(0, \sigma^2 I_d), \mathbb{P} * \mathcal{N}(0, \sigma^2 I_d))\right] = \omega\left(\frac{1}{n}\right).$$

Previous results. [GGNWP20] shows when $K < \sigma/2$, $\mathbb{E}\left[W_2^2(\mathbb{P}_n * \mathcal{N}(0, \sigma^2 I_d), \mathbb{P} * \mathcal{N}(0, \sigma^2 I_d))\right]$ converges with rate $\mathcal{O}\left(\frac{1}{n}\right)$; when $K > \sqrt{2}\sigma$, $\mathbb{E}\left[W_2^2(\mathbb{P}_n * \mathcal{N}(0, \sigma^2 I_d), \mathbb{P} * \mathcal{N}(0, \sigma^2 I_d))\right]$ converges with rate $\omega\left(\frac{1}{n}\right)$. Here is an obvious gap between $K < \sigma/2$ and $K > \sqrt{2}\sigma$, and our results close this gap. Moreover, [GGNWP20] shows that $\mathbb{E}\left[W_2(\mathbb{P}_n * \mathcal{N}(0, \sigma^2 I_d), \mathbb{P} * \mathcal{N}(0, \sigma^2 I_d))\right]$ converges with rate $\mathcal{O}\left(\frac{1}{n^{1/4}}\right)$ for any K and $\sigma > 0$.

Proof Idea. Let us introduce the χ^2-mutual information for a pair of random variables S, Y as

$$I_{\chi^2}(S; Y) \triangleq \chi^2(P_{S,Y} \| P_S \otimes P_Y),$$

where $\chi^2(P\|Q) = \int \left(\frac{dP}{dQ}\right)^2 dQ - 1$.

We will consider the case where $S \sim \mathbb{P}$, $Y = S + Z$ with $Z \sim \mathcal{N}(0, \sigma^2)$ independent to S. According to [GGNWP20], the convergence rate of smoothed empirical measure under W_2, KL-divergence and the

χ^2-divergence is closely related to $I_{\chi^2}(S;Y)$:
(**Proposition 6** in [GGNWP20]) If $\mathbb{P}$ is K-subgaussian with $K<\sigma$ and $I_{\chi^2}(S;Y)<\infty$, then

$$\mathbb{E}\left[W_2^2(\mathbb{P}_n * \mathcal{N}(0,\sigma^2 I_d), \mathbb{P} * \mathcal{N}(0,\sigma^2 I_d))\right] = \mathcal{O}\left(\frac{1}{n}\right).$$

(**Corollary 2** in [GGNWP20]) If $I_{\chi^2}(S;Y)=\infty$, then for any $\tau<\sigma$,

$$\mathbb{E}\left[W_2^2(\mathbb{P}_n * \mathcal{N}(0,\tau^2 I_d), \mathbb{P} * \mathcal{N}(0,\tau^2 I_d))\right] = \omega\left(\frac{1}{n}\right).$$

Hence our results follow from the following main technical propositions.

Proposition 1. *When $K<\sigma$, for any K-subgaussian d-dimensional distribution $\mathbb{P}$, we have $I_{\chi^2}(S;Y)<\infty$, where $S\sim\mathbb{P}, Z\sim\mathcal{N}(0,\sigma^2 I_d), S \perp\!\!\!\perp Z$ and $Y=S+Z$.*

Proposition 2. *When $K>\sigma$, there exists some 1-dimensional K-subgaussian distribution $\mathbb{P}$ such that $I_{\chi^2}(S;Y)=\infty$ for $S\sim\mathbb{P}, Z\sim\mathcal{N}(0,\sigma^2), S \perp\!\!\!\perp Z$ and $Y=S+Z$.*

We will prove these two propositions in the following two sections separately.

We note that results from [GGNWP20] and Proposition 1 also imply that $\mathbb{E}[D_{KL}(P_n * \mathcal{N}(0,\sigma^2 I_d) \| P * \mathcal{N}(0,\sigma^2 I_d))]$ and $\mathbb{E}[\chi^2(P_n * \mathcal{N}(0,\sigma^2 I_d) \| P * \mathcal{N}(0,\sigma^2 I_d))]$ both converge with rate $O\left(\frac{1}{n}\right)$. Our second Proposition 2 implies that for the special $\mathbb{P}$ constructed there we have

$$\mathbb{E}[D_{KL}(\mathbb{P}_n * \mathcal{N}(0,\sigma^2 I_d) \| \mathbb{P} * \mathcal{N}(0,\sigma^2 I_d)) = \omega\left(\frac{1}{n}\right)$$
$$\mathbb{E}\chi^2(\mathbb{P}_n * \mathcal{N}(0,\sigma^2 I_d) \| \mathbb{P} * \mathcal{N}(0,\sigma^2 I_d)) = \infty\,.$$

□

Next, we give a tight characterization on the W_2-convergence rate in dimension $d=1$.

Theorem 2. *[Improved bounds for dimension-1] Fix $K>\sigma>0$ and let*

$$\alpha = \frac{(\sigma^2+K^2)^2}{4(\sigma^4+K^4)}\,.$$

1. *(Lower Bound) With the choice of $\delta_n = \frac{1}{\sqrt[3]{\log\log n}}$, which converges to 0 as n goes to infinity, there exists some K-subgaussian distribution $\mathbb{P}$ over $\mathbb{R}$ such that*

$$\limsup_{n\to\infty} n^{\alpha+\delta_n}\mathbb{E}\left[W_2(\mathbb{P}_n * \mathcal{N}(0,\sigma^2 I_d), \mathbb{P} * \mathcal{N}(0,\sigma^2 I_d))\right] > 0\,.$$

2. *(Upper Bound) There exists a sequence $0<\delta_n\to 0$ such that for any 1-dimensional K-subgaussian $\mathbb{P}$ over $\mathbb{R}$ and $n\geq 2$, we have*

$$\mathbb{E}\left[W_2^2(\mathbb{P} * \mathcal{N}(0,\sigma^2), \mathbb{P}_n * \mathcal{N}(0,\sigma^2))\right] \leq n^{-2\alpha+\delta_n} \tag{7}$$

Remark 1. *With more work we believe that our proof gives $\delta_n = \frac{1}{\sqrt[3]{\log n}}$ in the upper bound part.*

Remark 2. *According to the Cauchy-Schwarz inequality, we have*

$$\mathbb{E}\left[W_2(\mathbb{P} * \mathcal{N}(0,\sigma^2), \mathbb{P}_n * \mathcal{N}(0,\sigma^2))\right] \leq \sqrt{\mathbb{E}\left[W_2^2(\mathbb{P} * \mathcal{N}(0,\sigma^2), \mathbb{P}_n * \mathcal{N}(0,\sigma^2))\right]}.$$

Therefore, the lower bound part in Theorem 2 indicates that for any K and $\epsilon > 0$, there exists some K-subgaussian distribution $\mathbb{P}$ and $\sigma > 0$ such that

$$\limsup_{n\to\infty} n^{2\alpha+2\delta_n}\mathbb{E}\left[W_2^2(\mathbb{P}_n * \mathcal{N}(0,\sigma^2 I_d), \mathbb{P} * \mathcal{N}(0,\sigma^2 I_d))\right] > 0. \tag{8}$$

and upper bound part in Theorem 2 indicates that

$$\mathbb{E}\left[W_2(\mathbb{P} * \mathcal{N}(0,\sigma^2), \mathbb{P}_n * \mathcal{N}(0,\sigma^2))\right] \le n^{-\alpha+\delta_n/2}.$$

Finally we provide an upper bound on the convergence of smoothed empirical measures under KL divergence:

Theorem 3. *Suppose $\mathbb{P}$ is a d-dimensional K-subgaussian distribution, then for any $\sigma > 0$, we have*

$$\mathbb{E}\left[D_{KL}\left(\mathbb{P}_n * \mathcal{N}(0,\sigma^2 I_d) \middle\| \mathbb{P} * \mathcal{N}(0,\sigma^2 I_d)\right)\right] = \mathcal{O}\left(\frac{(\log n)^d}{n}\right).$$

Remark 3. *From Proposition 1 and 2 and also results from [GGNWP20], we know that when $\sigma > K$, the convergence rate is $\mathcal{O}\left(\frac{1}{n}\right)$. From the above theorem, we know that when $\sigma \le K$, the convergence rate is between $\omega\left(\frac{1}{n}\right)$ and $\mathcal{O}\left(\frac{(\log n)^d}{n}\right)$. Hence at $K = \sigma$, the KL divergence also experiences a change to a non-parametric convergence rate, although with only a poly-logarithmic slow-down. As we discussed in Corollary 1 this precludes a general, finite logarithmic-Sobolev constant for a Gaussian mixture $\mathbb{P} * \mathcal{N}(0,\sigma^2)$ when $\sigma < K$.*

1.2 Organization of this Paper

In Section 2 we will present the proof of Proposition 1. In Section 3 we will present the proof of Proposition 2. The proof of the lower bound part and the upper bound part of Theorem 2 will be presented in Section 4 and 5. Finally in Section 6, we will present the proof of Theorem 3.

1.3 Notations

Throughout this paper, we use $*$ to denote convolutions of two random variables, *i.e.* for $X \sim \mathbb{P}, Y \sim \mathbb{Q}, X \perp\!\!\!\perp Y$, we have $X + Y \sim \mathbb{P} * \mathbb{Q}$; we use $\otimes$ to denote the product of two random variables's laws, *i.e.* for $X \sim \mathbb{P}, Y \sim \mathbb{Q}, X \perp\!\!\!\perp Y$, we have $(X, Y) \sim \mathbb{P} \otimes \mathbb{Q}$; we use $\circ$ to denote the composition between a Markov kernel $P_{Y|X}$ and a distribution P_X, *e.g.* for Y generated according to $P_{Y|X}$ with X's prior distribution P_X, then $Y \sim P_{Y|X} \circ P_X$.

Furthermore, we use $\mathbf{P}(E)$ to denote the probability of event E, $\mathbb{E}_{\mathbb{P}}[\cdot]$ to denote the expectation with respect to distribution $\mathbb{P}$. We use $A_n = \mathcal{O}(B_n), A_n = \Omega(B_n)$ to denote that $A_n \le CB_n$ and $A_n \ge CB_n$ for some positive constant C independent of n. We use $A = \tilde{O}(B)$ to denote that $A_n \le CB_n \cdot \log^l n$ for some positive constant C, l. We further use $\|\cdot\|_2$ to denote Euclidean norm, and use I_d to denote the $d \times d$ identity matrix.

We will use $\varphi_{\sigma^2}(\cdot)$ to denote the density of d-dimensional multivariate normal distribution $\mathcal{N}(0, \sigma^2 I_d)$. And for 1-dimensional distributions we use Φ_σ to denote the CDF of $\mathcal{N}(0, \sigma^2)$.

2 Proof of Proposition 1

In this section, we provide a proof of Proposition 1. The proof idea is to notice that we can write $I_{\chi^2}(S; Y)$ as $\mathbb{E}\left[\chi^2\left(\mathcal{N}(S,\sigma^2 I_d) \| \mathbb{E}\mathcal{N}(S,\sigma^2 I_d)\right)\right]$, then we decompose $\mathbb{R}^d$ into several cubes c_i with finite diameters, and we prove an upper bound of $\mathbb{E}\left[\mathbf{1}_{S\in c_i}\chi^2\left(\mathcal{N}(S,\sigma^2 I_d) \| \mathbb{E}\mathcal{N}(S,\sigma^2 I_d)\right)\right]$ for each i.

Proof. We suppose that the distribution $\mathbb{P}$ is d-dimensional. Then with $S \sim \mathbb{P}, Z \sim \mathcal{N}(0, \sigma^2 I_d), S \perp\!\!\!\perp Z$ and $Y = S + Z$, we have

$$
\begin{aligned}
I_{\chi^2}(S;Y) &= \mathbb{E}\left[\chi^2\left(\mathcal{N}(S,\sigma^2 I_d)\|\mathbb{E}\mathcal{N}(S,\sigma^2 I_d)\right)\right] && (9)\\
&= \mathbb{E}\left[\int_{\mathbb{R}^d} \frac{\varphi_{\sigma^2 I_d}(\mathbf{y}-S)^2}{\mathbb{E}\varphi_{\sigma^2 I_d}(\mathbf{y}-S)} d\mathbf{y} - 1\right] && (10)\\
&= (\sqrt{2\pi}\sigma)^{-d}\mathbb{E}\left[\int_{\mathbb{R}^d} \frac{\exp\left(-\|\mathbf{y}-S\|_2^2/\sigma^2\right)}{\rho(\mathbf{y})} d\mathbf{y}\right] - 1, && (11)
\end{aligned}
$$

where $S \sim \mathbb{P}$ and

$$
\rho(\mathbf{y}) = \mathbb{E}\exp\left(-\|\mathbf{y}-S\|_2^2/(2\sigma^2)\right). \tag{12}
$$

Let us decompose $\mathbb{R}^d = \bigcup_i c_i$ as a union of cubes of diameter 2. Since $K < \sigma$, we have $\frac{K}{\sigma} < 1$. Hence we can choose small $\delta, \delta' > 0$ such that

$$
\sqrt{\frac{1}{(1+\delta)^2(1+\delta')}} > \frac{K}{\sigma}.
$$

Notice that, due to the K-subgaussianity of S, we have [BLM13, p. 26]

$$
\mathbb{E}[\exp\left(\frac{(1+\delta')(1+\delta)^2}{2\sigma^2}\|S\|^2\right)] < \infty \tag{13}
$$

We will use the following lower bounds on $\rho(\mathbf{y})$:[2]

$$
\begin{aligned}
\rho(\mathbf{y}) &\gtrsim \exp\left(-\frac{1+\delta'}{2\sigma^2}\|\mathbf{y}\|^2\right), && (14)\\
\rho(\mathbf{y}) &\gtrsim \mathbf{P}[S \in c_i]\exp\left(-\frac{3}{4\sigma^2}\|\mathbf{y}-s\|^2\right) \qquad \forall s \in c_i. && (15)
\end{aligned}
$$

Assuming these inequalities, the proof proceeds as follows. Fix an arbitrary $s \in \mathbb{R}^d$ and notice that from (14) whenever $\|\mathbf{y}\| \le (1+\delta)\|s\|$ we have

$$
\rho(\mathbf{y}) \gtrsim \exp\left(-\frac{(1+\delta')(1+\delta)^2}{2\sigma^2}\|s\|^2\right)
$$

which implies that

$$
\int_{\|\mathbf{y}\|\le(1+\delta)\|s\|} \frac{\exp\left(-\|\mathbf{y}-s\|_2^2/\sigma^2\right)}{\rho(\mathbf{y})} d\mathbf{y} \lesssim \exp\left(\frac{(1+\delta')(1+\delta)^2}{2\sigma^2}\|s\|^2\right), \tag{16}
$$

since the numerator integrates over $\mathbb{R}^d$ to $(\pi\sigma^2)^{d/2}$. On the other hand, from (15) if $s \in c_i$ then

$$
\frac{\exp\left(-\|\mathbf{y}-s\|_2^2/\sigma^2\right)}{\rho(\mathbf{y})} \lesssim \mathbf{P}[S \in c_i]^{-1}\exp\left(-\frac{\|\mathbf{y}-s\|^2}{4\sigma^2}\right). \tag{17}
$$

Note also that when $\|\mathbf{y}\| \ge (1+\delta)\|s\|$ we have $\|\mathbf{y}-s\| \ge \delta\|s\|$. Thus, integrating the right-hand side of (17) over $\{\mathbf{y} : \|\mathbf{y}-s\| \ge \delta\|s\|\}$ we obtain

$$
\mathbf{P}[S \in c_i]^{-1}\int_{\|\mathbf{y}-s\|\ge\delta\|s\|} \exp\left(-\frac{\|\mathbf{y}-s\|^2}{4\sigma^2}\right) \lesssim \mathbf{P}[S \in c_i]^{-1}\mathbf{P}\left[U_d > \frac{\delta^2\|s\|^2}{\sqrt{2\sigma^2}}\right],
$$

[2] Notation $\gtrsim$ and $\lesssim$ in this proof denote inequalities up to constants that may depend on K, σ, d and distribution P.

where U_d denotes a $\chi^2(d)$ random variable with d degrees of freedom. Using Chernoff inequality $\mathbf{P}[U_d > r] \le 2^{\frac{d}{2}} e^{-r/4}$ we obtain

$$\max_{s \in c_i} \mathbf{P}\left[U_d > \frac{\delta^2 \|s\|^2}{\sqrt{2\sigma^2}}\right] \lesssim \exp(-C\|x_i\|^2)\,,$$

where x_i is the center of the cube c_i and C is some constant.

Thus, together with (16) we obtain that for any $s \in c_i$:

$$\chi^2(\mathcal{N}(s, \sigma^2 I_d)\|P_Y) \lesssim \int_{\mathbb{R}^d} \frac{\exp\left(-\|\mathbf{y}-s\|_2^2/\sigma^2\right)}{\rho(\mathbf{y})} d\mathbf{y} \lesssim \mathbf{P}[S \in c_i]^{-1} \exp(-C\|x_i\|^2) + \exp\left(\frac{(1+\delta')(1+\delta)^2}{2\sigma^2}\|s\|^2\right).$$

Taking expectation of the latter over S, the second term is finite because of (13), while the first one is bounded because the number of cubes with $\|x_i\| \le r$ is $O(r^d)$. This completes the proof of finiteness of (9), assuming (14) and (15). We now establish these.

To show (14) set t to be any value such that $\mathbf{P}[\|S\| < t] \ge \frac{1}{2}$ and notice that

$$\rho(\mathbf{y}) \gtrsim \mathbb{E}[\exp\left(-\frac{\|\mathbf{y}-S\|^2}{2\sigma^2}\right) | \|S\| < t]\,. \tag{18}$$

Next, notice that for any t and $\delta' > 0$ we can find some constant C' such that

$$(a+t)^2 \le (1+\delta')a^2 + C', \qquad \forall a \in \mathbb{R}\,. \tag{19}$$

Thus for any $\|s\| < t$ we have

$$\exp\left(-\frac{\|\mathbf{y}-s\|^2}{2\sigma^2}\right) \ge \exp\left(-\frac{(\|\mathbf{y}\|+\|s\|)^2}{2\sigma^2}\right) \gtrsim \exp\left(-\frac{\|\mathbf{y}\|^2(1+\delta')}{2\sigma^2}\right).$$

Using this estimate in (18) recovers (14).

To show (15) we start similarly:

$$\rho(\mathbf{y}) \ge \mathbf{P}[S \in c_i]\mathbb{E}[\exp\left(-\frac{\|\mathbf{y}-S\|^2}{2\sigma^2}\right) | S \in c_i] \tag{20}$$

Now fix any (non-random) $s \in c_i$ and notice that under the conditioning above we have $\|S - s\| \le 2$ because the cube c_i has diameter 2. Then from triange inequality and (19) with $\delta' = 1/2$ we obtain

$$\|\mathbf{y}-S\|^2 \le (\|\mathbf{y}-s\|+2)^2 \le \frac{3}{2}\|\mathbf{y}-s\|^2 + C''\,.$$

Using this bound in (20) yields (15). □

For future reference we also need to show that the Rényi mutual information $I_\lambda(S;Y)$ is also finite for all $1 < \lambda < 2$. The Rényi mutual information is defined as follows:

Definition 1 (Rényi Divergence and Rényi Mutual Information [Rén61])**.** *Assume random variables (X, Y) have joint distribution $P_{X,Y}$. For any $\lambda > 1$, the Rényi divergence and Rényi Mutual Information of order λ are defined as*

$$I_\lambda(X;Y) \triangleq D_\lambda(P_{X,Y}\|P_X \otimes P_Y),$$

where we use P_X, P_Y to denote the marginal distribution with respect to X and Y, and $P_X \otimes P_Y$ denotes the joint distribution of (X', Y') where $X' \sim P_X, Y' \sim P_Y$ are independent to each other, and the Rényi divergence between two distributions $\mathbb{P}$ and $\mathbb{Q}$ is defined as $D_\lambda(\mathbb{P}\|\mathbb{Q}) \triangleq \frac{1}{\lambda-1}\log \mathbb{E}_{\mathbb{Q}}\left[\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)^\lambda\right]$.

We summarize the result below.

Lemma 1. *Suppose random variables $S \sim \mathbb{P}, Z \sim \mathcal{N}(0, \sigma^2 I_d)$ are independent of each other. Let $Y = S + Z$. Fix $1 < \lambda < 2$ and let $l = l(\lambda) = \frac{\lambda-1}{2-\lambda}(d+1)$. If the random variable $S \sim \mathbb{P}$ has finite moment l-th moment $\mathcal{M} \triangleq \mathbb{E}[\|S\|_2^l]^{2-\lambda}$, then for any $\sigma > 0$ there exists a constant $C = C(\mathbb{P}, \sigma) > 0$ such that:*

$$I_\lambda(S; Y) \leq \frac{1}{\lambda - 1} \log (C\mathcal{M}) \,.$$

Moreover, if $\mathbb{P}$ is a K-subgaussian distribution, we have for all $1 < \lambda < 2$

$$I_\lambda(S; Y) \leq \frac{1}{\lambda - 1} \log \left(\frac{C'}{(2-\lambda)^d} \right)$$

for some constant $C' = C(\mathbb{P}, K, \sigma) > 0$.

Proof. According to the definition of Rényi divergence, we have

$$I_\lambda(S; Y) = \frac{1}{\lambda - 1} \log \left(C_0 \mathbb{E} \left[\int \frac{\rho^\lambda_{Y|S}(\mathbf{y}|S)}{\rho_Y(\mathbf{y})^{\lambda-1}} d\mathbf{y} \right] \right),$$

for some positive constant C_0, where $\rho_{Y|S}(\mathbf{y}|S) = \exp(-\frac{\|\mathbf{y}-S\|^2}{2\sigma^2})$ and $\rho_Y(\mathbf{y})$ is from (12) Therefore, we only need to prove

$$\mathbb{E} \left[\int \frac{\rho^\lambda_{Y|S}(\mathbf{y}|S)}{\rho_Y(\mathbf{y})^{\lambda-1}} d\mathbf{y} \right] \lesssim \mathcal{M}$$

for distributions $\mathbb{P}$ with finite l-th moment, and

$$\mathbb{E} \left[\int \frac{\rho^\lambda_{Y|S}(\mathbf{y}|S)}{\rho_Y(\mathbf{y})^{\lambda-1}} d\mathbf{y} \right] \lesssim \frac{1}{(2-\lambda)^d}$$

for K-subgaussian distribution $\mathbb{P}$.

We write $\mathbb{R} = \bigcup_i c_i$ as a union of cubes of diameter 2. For any $s \in c_i$, we have

$$\rho_Y(\mathbf{y}) = \mathbb{E} \left[\exp \left(-\frac{\|\mathbf{y} - S\|^2}{2\sigma^2} \right) \right] \geq \mathbf{P}[S \in c_i] \mathbb{E} \left[\exp \left(-\frac{\|\mathbf{y} - S\|^2}{2\sigma^2} \right) \middle| S \in c_i \right].$$

We further notice that for any $S, s \in c_i$, we have $\|S - s\| \leq 2$, implying

$$\exp \left(-\frac{\|\mathbf{y} - S\|^2}{2\sigma^2} \right) \geq \exp \left(-\frac{3\|\mathbf{y} - s\|^2}{4\sigma^2} - \frac{12}{2\sigma^2} \right) = \exp \left(-\frac{6}{\sigma^2} \right) \exp \left(-\frac{3\|\mathbf{y} - s\|^2}{4\sigma^2} \right)$$

following from inequality

$$\|\mathbf{y} - S\|^2 \leq (\|\mathbf{y} - s\| + 2)^2 = \|\mathbf{y} - s\|^2 + 4\|\mathbf{y} - s\| + 4 \leq \frac{3}{2}\|\mathbf{y} - s\|^2 + 12 \,.$$

Hence we obtain that

$$\rho_Y(\mathbf{y}) \geq \exp \left(-\frac{6}{\sigma^2} \right) \mathbf{P}[S \in c_i] \exp \left(-\frac{3\|\mathbf{y} - s\|^2}{4\sigma^2} \right),$$

which indicates that

$$\begin{aligned} \frac{\exp\left(-\lambda\|\mathbf{y} - s\|_2^2/(2\sigma^2)\right)}{\rho_Y(\mathbf{y})^{\lambda-1}} &\leq \exp \left(\frac{6(\lambda-1)}{\sigma^2} \right) \mathbf{P}[S \in c_i]^{1-\lambda} \exp \left(-\frac{(3-\lambda)\|\mathbf{y} - s\|^2}{4\sigma^2} \right) \\ &\leq \exp \left(\frac{6}{\sigma^2} \right) \mathbf{P}[S \in c_i]^{1-\lambda} \exp \left(-\frac{\|\mathbf{y} - s\|^2}{4\sigma^2} \right) \end{aligned}$$

after noticing the fact that $1 \le \lambda < 2$. Therefore for any $s \in \mathbb{R}^d$ we have

$$\int_{\mathbb{R}^d} \frac{\exp\left(-\lambda\|\mathbf{y}-s\|_2^2/(2\sigma^2)\right)}{\rho_Y(\mathbf{y})^{\lambda-1}} d\mathbf{y} \lesssim \mathbf{P}[S \in c_i]^{1-\lambda} \int_{\mathbb{R}^d} \exp\left(-\frac{\|\mathbf{u}\|^2}{4\sigma^2}\right) d\mathbf{u} \lesssim \mathbf{P}[S \in c_i]^{1-\lambda},$$

where we use $\lesssim$ to hide constant factors depending on σ, d. Taking the expectation over S, we obtain that

$$\mathbb{E}\left[\frac{\rho_{Y|S}^{\lambda}(\mathbf{y}|S)}{\rho_Y(\mathbf{y})^{\lambda-1}}\right] \lesssim \sum_i \mathbf{P}[S \in c_i]^{2-\lambda}. \tag{21}$$

Next, we use L_r to denote the set of cubes whose centers belong to $\{r-1 \le \|x_i\| < r\}$. Then we have $|L_r| = \mathcal{O}(r^{d-1})$. We further let $p_r = \sum_{c_i \in M_r} \mathbf{P}[S \in c_i]$, then according to Jensen's inequality we obtain that

$$\sum_{c_i \in L_r} \mathbf{P}[S \in c_i]^{2-\lambda} \le L_r \cdot \left(\frac{1}{L_r}\sum_{c_i \in M_r} \mathbf{P}[S \in c_i]\right)^{2-\lambda} = L_r \cdot \left(\frac{p_r}{L_r}\right)^{2-\lambda} = L_r^{\lambda-1} p_r^{2-\lambda}.$$

Assuming $\mathcal{M}_{\frac{(\lambda-1)(d+1)}{2-\lambda}} < \infty$, we have for any $1 < \lambda < 2$,

$$\begin{aligned}
\sum_{i=1} \mathbf{P}[S \in c_i]^{2-\lambda} &= \sum_{r=1}^{\infty} \sum_{c_i \in M_r} \mathbf{P}[S \in c_i] \lesssim \sum_{r=1}^{\infty} r^{(\lambda-1)(d-1)} p_r^{2-\lambda} \\
&\le \left(\sum_{r=1}^{\infty} r^{\frac{(\lambda-1)(d-1)+2(\lambda-1)}{2-\lambda}} p_r\right)^{2-\lambda} \left(\sum_{r=1}^{\infty} \frac{1}{r^2}\right)^{\lambda-1} \lesssim \mathbb{E}[\|S\|_2^l]^{2-\lambda} = \mathcal{M}
\end{aligned}$$

where in the second last inequality we use the Hölder inequality. As for the K-subgaussian cases, we notice that $p_r \lesssim \exp\left(-\frac{r^2}{2K^2}\right)$. Therefore, we obtain that

$$\begin{aligned}
&\sum_{i=1} \mathbf{P}[S \in c_i]^{2-\lambda} = \sum_{r=1}^{\infty} \sum_{c_i \in M_r} \mathbf{P}[S \in c_i]^{2-\lambda} \le \sum_{r=1}^{\infty} |M_r| p_r^{2-\lambda} \lesssim \sum_{r=1}^{\infty} r^{d-1} \exp\left(-\frac{(2-\lambda)r^2}{2K^2}\right) \\
&\le \sum_{r=1}^{\infty} r^{d-1} \exp\left(-\frac{(2-\lambda)r}{2K^2}\right) \le d! \left(1 - \exp\left(-\frac{2-\lambda}{2K^2}\right)\right)^{-d} \lesssim \frac{1}{(2-\lambda)^d},
\end{aligned}$$

where we use the fact that $\sum_{k=0}^{\infty} (k+1)^d c^{-k} \le d! c^{-k-1}$ for any $0 < c < 1$, and $1 - \exp(-x) \le 1 - (1-x) = x$ holds for any $x \in \mathbb{R}$.

Based on these two upper bounds on $\sum \mathbf{P}[S \in c_i]^{2-\lambda}$, (21) yields the desired bounds on $I_\lambda(S;Y)$. $\square$

3 Proof of Proposition 2

In this section, we will present a proof of Proposition 2. The main idea of this proof is to construct a hard example $\mathbb{P} = \sum_{k=0}^{\infty} p_k \delta_{r_k}$ with subgaussian tails, where r_i and r_j are far away from each other so that $\delta_{r_j} * \mathcal{N}(0, \sigma^2)$ with $j \ne i$ has very little effects on the density of $\mathbb{P} * \mathcal{N}(0, \sigma^2)$ near r_i. Therefore we can uniformly lower bound $\mathbb{E}\left[\mathbf{1}_{S=r_i} \chi^2\left(\mathcal{N}(S, \sigma^2 I_d) \| \mathbb{E}\mathcal{N}(S, \sigma^2 I_d)\right)\right]$ for each i, and if we sum up over all i we can prove the infiniteness of $I_{\chi^2}(S;Y)$.

Proof. Without loss of generality we assume $\sigma = 1$, and we only need to prove the proposition for $K > 1$. (Otherwise we consider $S' = S/\sigma, Z' = Z/\sigma$ and $Y' = Y/\sigma$, and we will have S' is a K/σ-subgaussian distribution, $Z' \sim \mathcal{N}(0,1)$ and $I_{\chi^2}(S;Y) = I_{\chi^2}(S';Y')$.)

We construct a 1-dimensional distribution $\mathbb{P}$ similarly to [GGNWP20] as follows:

$$\mathbb{P} = \sum_{k=0}^{\infty} p_k \delta_{r_k},$$

where we choose $r_0 = 0$, $p_0 = 1 - \sum_{k=1}^{\infty} p_k$ and for some positive constant c_1 to be determined we choose

$$p_k = c_1 \exp\left(-\frac{r_k^2}{2K^2}\right), \quad k \geq 1. \tag{22}$$

Here we let r_i be a geometric sequence:

$$r_1 = 1, \quad r_{i+1} = cr_i, \ \forall i \geq 1,$$

where $c > 2$ is a constant to be specified later. We restrict that c_1 only depends on K and

$$c_1 \cdot \sum_{k=1}^{\infty} \exp\left(-\frac{r_k^2}{2K^2}\right) < 1.$$

Then we will have $p_0 = 1 - \sum_{k=1}^{\infty} p_k > 0$ making $\mathbb{P}$ a well-defined distribution. In Appendix A we show that there exists a $c_1 > 0$ such that for any constant $c > 2$ the distribution $\mathbb{P}$ is K-subgaussian.

In the following, we establish a weaker claim that $S \sim \mathbb{P}$ satisfies

$$\forall \alpha : \qquad \mathbb{E}[\exp(\alpha S)] \leq 2 \exp\left(\frac{\alpha^2 K^2}{2}\right).$$

Note that this is slightly weaker than the definition of K-subgaussianity (2). Indeed, we notice that

$$\mathbb{E}[\exp(\alpha S)] = p_0 + c_1 \sum_{k=1}^{\infty} \exp\left(-\frac{1}{2K^2}\left(r_k - \alpha K^2\right)^2\right) \exp\left(\frac{K^2\alpha^2}{2}\right).$$

We suppose k_0 to be the smallest k such that $r_k - \alpha K^2$ to be positive. Since $r_{k+1} - r_k \geq 1$ for every k, we have for $k \geq k_0$, $r_k - \alpha K^2 \geq k - k_0 + r_{k_0} - \alpha K^2 \geq k - k_0$, and for $k < k_0$, $r_k - \alpha K^2 \leq r_{k_0 - 1} - \alpha K - (k_0 - 1 - k) \leq -(k_0 - 1 - k)$ since $r_{k_0-1} - \alpha K^2 \leq 0$. Hence, we have

$$\begin{aligned}
\sum_{k<k_0} + \sum_{k \geq k_0} e^{-\frac{1}{2K^2}\left(r_k - \alpha K^2\right)^2} &\leq \sum_{k<k_0} e^{-\frac{(k_0-1-k)^2}{2K^2}} + \sum_{k \geq k_0} e^{-\frac{(k-k_0)^2}{2K^2}} \\
&\leq 2\sum_{k=0}^{\infty} e^{-\frac{k}{2K^2}} = \frac{2}{1 - \exp\left(-\frac{1}{2K^2}\right)}.
\end{aligned}$$

Therefore, if we choose $c_1 = \frac{1 - \exp\left(-\frac{1}{2K^2}\right)}{2}$, and notice that $p_0 \leq 1 \leq \exp\left(\frac{K^2\alpha^2}{2}\right)$, we would have

$$\mathbb{E}[\exp(\alpha S)] \leq \exp\left(\frac{K^2\alpha^2}{2}\right) + \exp\left(\frac{K^2\alpha^2}{2}\right) = 2\exp\left(\frac{K^2\alpha^2}{2}\right).$$

For now we proceed assuming that c_1 is already chosen such that for any $c > 2$, we have that $\mathbb{P}$ is a K-subgaussian distribution. Then, our goal is to specify a value of constant c such that $I_{\chi^2}(S;Y) = \infty$.

From the definition of I_{χ^2}, we have the following chain:

$$
\begin{aligned}
I_{\chi^2}(S;Y) &\triangleq \int_{\mathbb{R}} \frac{\mathbb{E}\varphi_1^2(y-S)}{\mathbb{E}\varphi_1(y-S)} dy - 1 \\
&= \int_{\mathbb{R}} \frac{\sum_{k=0}^{\infty} p_k \varphi_{\frac{1}{\sqrt{2}}}(y-r_k)}{\sum_{k=1}^{\infty} p_k \varphi_1(y-r_k)} dy \\
&= \sum_{k=0}^{\infty} \int_{\mathbb{R}} \frac{\varphi_{\frac{1}{\sqrt{2}}}(y-r_k)}{\varphi_1(y-r_k)} \cdot \frac{1}{1+\sum_{j\neq k} \frac{p_j}{p_k} \frac{\varphi_1(y-r_j)}{\varphi_1(y-r_k)}} dy \\
&\geq \sum_{k=0}^{\infty} \int_{r_k-\delta}^{r_k+\delta} \frac{\varphi_{\frac{1}{\sqrt{2}}}(y-r_k)}{\varphi_1(y-r_k)} \cdot \frac{1}{1+\sum_{j\neq k} \frac{p_j}{p_k} \frac{\varphi_1(y-r_j)}{\varphi_1(y-r_k)}} dy\,.
\end{aligned} \tag{23}
$$

where we fixed arbitary $\delta > 0$. Below, we will show that for some $\delta \in (0,1)$ and $C' > 0$ we have for all k and $|y - r_k| < \delta$:

$$
1 + \sum_{j\neq k} \frac{p_j}{p_k} \frac{\varphi_1(y-r_j)}{\varphi_1(y-r_k)} \leq C + \sum_{j=1, j\neq k}^{\infty} \exp(-j/2) < C'. \tag{24}
$$

Assuming this, we continue (23) as follows:

$$
\geq \frac{1}{C'} \sum_{k=0}^{\infty} \int_{r_k-\delta}^{r_k+\delta} \frac{\varphi_{\frac{1}{\sqrt{2}}}(y-r_k)}{\varphi_1(y-r_k)} dy = \geq \frac{1}{C'} \sum_{k=0}^{\infty} \int_{-\delta}^{+\delta} \frac{\varphi_{\frac{1}{\sqrt{2}}}(y)}{\varphi_1(y)} dy = \infty\,.
$$

To show (24) we first consider $j = 0$ and $|z - r_k| \leq \delta$, we have

$$
\begin{aligned}
&\frac{p_j}{p_k} \frac{\varphi_1(y-r_j)}{\varphi_1(y-r_k)} \leq \frac{\varphi_1(y)}{p_k \varphi_1(y-r_k)} \leq \frac{1}{c_1} \exp\left(-\frac{y^2}{2} + \frac{r_k^2}{2K^2} + \frac{(y-r_k)^2}{2}\right) \\
&\leq \frac{1}{c_1} \exp\left(-\frac{(r_k-\delta)^2}{2} + \frac{r_k^2}{2K^2} + \frac{\delta^2}{2}\right) \\
&\leq \frac{1}{c_1} \exp\left(-\frac{(r_k-\delta)^2}{2} + \frac{\left((r_k-\delta)^2 + \frac{K^2\delta^2}{1-K^2}\right)\left(1 + \frac{1-K^2}{K^2}\right)}{2K^2} + \frac{\delta^2}{2}\right) \\
&= \frac{1}{c_1} \exp\left(\frac{K^2\delta^2}{2(1-K^2)} + \frac{\delta^2}{2}\right) \triangleq C.
\end{aligned}
$$

For $j \geq 1$ and $|y - r_k| \leq \delta$, we have by bounding $y(r_j - r_k) \leq -r_k^2 + r_k r_j + \delta |r_k - r_j|$ the following chain

$$
\begin{aligned}
&\frac{p_j}{p_k} \frac{\varphi_1(y-r_j)}{\varphi_1(y-r_k)} = \exp\left(\left(\frac{1}{2K^2} + \frac{1}{2}\right)(r_k^2 - r_j^2) - y(r_k - r_j)\right) \\
&\leq \exp\left(\left(\frac{1}{2K^2} + \frac{1}{2}\right)(r_k^2 - r_j^2) - r_k^2 + r_k r_j + \delta |r_k - r_j|\right) \\
&\leq \exp\left(\left(\frac{1}{2K^2} + \frac{1}{2} - 1\right) r_k^2 - \left(\frac{1}{2K^2} + \frac{1}{2}\right) r_j^2 + r_k r_j + \delta r_k + \delta r_j\right) \\
&= \exp(A + B + C - r_j^2/4)
\end{aligned}
$$

where we denoted

$$
\begin{aligned}
A &\triangleq \frac{l}{2} r_k^2 - \frac{1}{2K^2} r_j^2 + r_k r_j & \qquad \ell \triangleq \frac{1}{2K^2} - \frac{1}{2} \\
B &\triangleq \frac{\ell}{2} r_k^2 + \delta r_k \\
C &\triangleq -\frac{1}{4} r_j^2 + \delta r_j\,.
\end{aligned}
$$

Note that $K > 1$ and, thus, $\ell < 0$. We show that by choosing c and δ it is possible to make sure $A, B, C \le 0$ for all k, j. First, notice that because $r_k \ge 1$ or $r_k = 0$ by setting $\delta = \min\left(-\frac{\ell}{2}, \frac{1}{4}\right)$ we have $B, C \le 0$.

Second, we have $A = r_j^2 f(r_k/r_j)$ where $f(y) = \frac{\ell}{2}y^2 + y - \frac{1}{2K^2}$. Since $f(0) < 0$ and $f(+\infty) = -\infty$ (recall $\ell < 0$) we must have that for some sufficiently large $c > 0$ we have $f(y) < 0$ if $y \le 1/c$ or $y \ge c$. For convenience we take this $c > 2$ as well. Since r_k/r_j is always either $\le 1/c$ or $\ge c$ we conclude $A \le 0$.

Continuing, we obtained that with our choice of c, for $j \ne k, j \ge 1$ and $|y - r_k| \le \delta$ we have

$$\begin{aligned}\frac{p_j}{p_k}\frac{\varphi_1(y-r_j)}{\varphi_1(y-r_k)} &\le \exp\left(A + B + C - \frac{r_j^2}{4}\right) \le \exp\left(-\frac{r_j^2}{4}\right)\\ &\le \exp\left(-\frac{c^j}{4}\right) \le \exp\left(-\frac{2^j}{4}\right) \le \exp(-j/2),\end{aligned}$$

which indicates that $\exists C'$ such that (24) holds. Therefore,

$$\begin{aligned}&\sum_{k=0}^{\infty}\int_{\mathbb{R}}\frac{\varphi_{\frac{1}{\sqrt{2}}}(y-r_k)}{\varphi_1(y-r_k)}\cdot\frac{1}{1+\sum_{j\ne k}\frac{p_j}{p_k}\frac{\varphi_1(y-r_j)}{\varphi_1(y-r_k)}}dy\\ &\ge \sum_{k=0}^{\infty}\int_{r_k-\delta}^{r_k+\delta}\frac{\varphi_{\frac{1}{\sqrt{2}}}(y-r_k)}{\varphi_1(y-r_k)}\cdot\frac{1}{1+\sum_{j\ne k}\frac{p_j}{p_k}\frac{\varphi_1(y-r_j)}{\varphi_1(y-r_k)}}dy\\ &\ge \left(\int_{-\delta}^{\delta}\frac{\varphi_{\frac{1}{\sqrt{2}}}(y)}{\varphi_1(y)}dy\right)\cdot\sum_{k=0}^{\infty}\frac{1}{C'}\\ &= \infty\end{aligned}$$

And we have proved that $I_{\chi^2}(S; Y) = \infty$. $\square$

4 Proof of the Lower Bound in Theorem 2

To begin with, we consider a simple Bernoulli distribution case, which shares lots properties in common with the counter example we construct in order to prove the lower bound of Theorem 2.

4.1 A Warmup Example: Simple Bernoulli Distribution Case

We consider a Bernoulli distribution $\mathbb{P}_h = (1 - p_h)\delta_0 + p_h\delta_h$ with $p_h = \exp\left(-\frac{h^2}{2K^2}\right)$. The behavior of the lower bound of

$$\sup_h \mathbb{E}\left[W_2(\mathbb{P}_h * \mathcal{N}(0, \sigma^2), \mathbb{P}_{h,n} * \mathcal{N}(0, \sigma^2))\right]$$

shares the same rate as the lower bound in Theorem 2.

Proposition 3. *For some $h > 0$, we define $\mathbb{P}_h = (1-p)\delta_0 + p\delta_h$, with $p = e^{-h^2/(2K^2)}$, then for any $K, \sigma > 0$ and $\epsilon > 0$,*

$$\sup_h \mathbb{E}\left[W_2(\mathbb{P}_h * \mathcal{N}(0, \sigma^2), \mathbb{P}_{h,n} * \mathcal{N}(0, \sigma^2))\right] = \Omega\left(n^{-\alpha-\epsilon}\right),$$

where $\mathbb{P}_{h,n}$ is the empirical measure constructed from n i.i.d. samples from $\mathbb{P}_h$.

Our proof will rely on the following auxiliary lemma.

Lemma 2. *Suppose two 1-dimensional distribution μ, ν with CDFs F_μ, F_ν satisfy $F_\mu(t) \ge F_\nu(t + 2)$, then we have*

$$W_2(\mu, \nu)^2 \ge \mathbf{P}(Y \in [t + 1, t + 2]), \quad Y \sim \nu.$$

Proof. We consider the optimal coupling between (X, Y), then the optimal coupling is the quantile-quantile coupling since μ, ν are 1-dimensional distributions. Noticing that $F_\mu(t) \geq F_\nu(t+2)$, all mass between $[t+1, t+2]$ in ν will transport to places below t. Therefore, we have

$$W_2(\mu, \nu)^2 \geq \mathbf{P}(Y \in [t+1, t+2]).$$

$\square$

Proof of Proposition 3. Given $h > 0$, we assume $\mathbb{P}_{h,n} = (1 - \hat{p}_h)\delta_0 + \hat{p}_h \delta_h$, where $\hat{p}_h = \frac{1}{n}\left(\sum_{k=1}^n \mathbf{1}_{X_k = h}\right)$, and $X_1, \cdots, X_n \sim \mathbb{P}_h$ are i.i.d. In the following proof, when there is no danger of confusion, we use $\tilde{F}_{n,\sigma}, F_\sigma, \Phi_\sigma$ to denote the CDF of $\mathbb{P}_{h,n} * \mathcal{N}(0, \sigma^2), \mathbb{P}_h * \mathcal{N}(0, \sigma^2), \mathcal{N}(0, \sigma^2)$. We will prove the results for ϵ sufficiently small, since cases of larger ϵ are direct corollary of cases of small ϵ.

We fix σ, K, and let $\delta = \delta(\sigma, K, \epsilon)$ such that

$$\frac{(1+\delta)(1+\sigma^2/K^2)^2}{2(1-\delta)(1+\sigma^2/K^2) - 4\delta\sigma^2/K^2} = \frac{(1+\sigma^2/K^2)^2}{2 + 2\sigma^2/K^2} + 2\epsilon, \tag{25}$$

and we let $\zeta = \frac{\left(\frac{1}{2} + \frac{\sigma^2}{2K^2}\right)^2}{2\sigma^2}$. Then we know that

$$\lim_{\epsilon \to 0} \delta(\sigma, K, \epsilon) = 0,$$

and $\zeta - \frac{1}{2K^2} = \frac{\left(\frac{1}{2} - \frac{\sigma^2}{2K^2}\right)^2}{2\sigma^2} > 0$ With loss of generality we assume $\delta < \frac{1}{2}$. Therefore, for sufficiently small $\epsilon > 0$, we will have $\delta = \delta(\sigma, K, \epsilon) < \min\left\{\frac{1}{2}, 1 - \frac{1}{2K^2\zeta}\right\}$.

We first show that for sufficiently large h and some specific choice of $t \in (0, h-2)$, we will have

$$\begin{aligned}
\mathbf{P}(X \in [t, t+2]) &\leq \frac{4}{\sqrt{2\pi}\sigma} \exp\left(-(1-\delta)\zeta h^2\right) \\
\mathbf{P}(X \in [t+1, t+2]) &\geq \frac{1}{2\sqrt{2\pi}\sigma} \exp\left(-(1+\delta)\zeta h^2\right).
\end{aligned}$$

Actually, we have the following estimation of the probability of $\mathbb{P}_h * \mathcal{N}(0, \sigma^2)$ within the intervals $[t, t+2]$ and $[t+1, t+2]$: for $X \sim \mathbb{P}_h * \mathcal{N}(0, \sigma^2)$ and $t \in (0, h-2)$, we have

$$\begin{aligned}
\mathbf{P}(X \in [t, t+2]) &\leq 2 \cdot \max_{t' \in [t, t+2]} \left[\frac{1-p_h}{\sqrt{2\pi}\sigma} \exp\left(-\frac{t'^2}{2\sigma^2}\right) + \frac{p_h}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(h-t')^2}{2\sigma^2}\right)\right] &(26) \\
&\leq \frac{2}{\sqrt{2\pi}\sigma} \cdot \left[\exp\left(-\frac{t^2}{2\sigma^2}\right) + \exp\left(-\frac{h^2}{2K^2} - \frac{(h-t-2)^2}{2\sigma^2}\right)\right], &(27) \\
\mathbf{P}(X \in [t+1, t+2]) &\geq \min_{t' \in [t+1, t+2]} \left[\frac{1-p_h}{\sqrt{2\pi}\sigma} \exp\left(-\frac{t'^2}{2\sigma^2}\right) + \frac{p_h}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(h-t')^2}{2\sigma^2}\right)\right] &(28) \\
&\geq \frac{1}{2\sqrt{2\pi}\sigma} \exp\left(-\frac{(t+2)^2}{2\sigma^2}\right), &(29)
\end{aligned}$$

where we have use the fact that $1 - p_h \geq \frac{1}{2}$ for all $h \geq 2K$. Next, we would like to select the value of t such that

$$-\frac{h^2}{2K^2} - \frac{(h-t)^2}{2\sigma^2} = -\frac{t^2}{2\sigma^2}.$$

That is,

$$t = \frac{h}{2} + \frac{\sigma^2 h}{2K^2}.$$

Since $\sigma < K$, we notice that $\exists \bar{h} > 0$ depending on σ, K such that for $h > \bar{h}$, we have $t \in (0, h-2)$, and when h goes to infinity, both t and $h-t$ go to infinity as well. Hence for any $0 < \delta < 1$ there exists C_1, C_h only depending on K, σ and δ such that when $h > C_h$, we have

$$\frac{(h-t-2)^2}{2\sigma^2} \le \frac{(1-\delta)(h-t)^2}{2\sigma^2} \quad \text{and} \quad \frac{(t+2)^2}{2\sigma^2} \le \frac{(1+\delta)t^2}{2\sigma^2},$$

which indicates that

$$\mathbf{P}(X \in [t, t+2]) \le \frac{4}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(1-\delta)t^2}{2\sigma^2}\right) = \frac{4}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(1-\delta)\left(\frac{1}{2} + \frac{\sigma^2}{2K^2}\right)^2 h^2}{2\sigma^2}\right), \tag{30}$$

$$\mathbf{P}(X \in [t+1, t+2]) \ge \frac{1}{2\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(1+\delta)t^2}{2\sigma^2}\right) = \frac{1}{2\sqrt{2\pi}\sigma} \exp\left(-\frac{(1+\delta)\left(\frac{1}{2} + \frac{\sigma^2}{2K^2}\right)^2 h^2}{2\sigma^2}\right) \tag{31}$$

holds for all $h > C_h$. We let $C_1 \triangleq \frac{4}{\sqrt{2\pi}\sigma}$.

We next show that fix $h > 0$, for any n such that $np_h \ge 128$ and $0 < t < h-2$, we have with probability at least $\frac{1}{16}$,

$$\tilde{F}_{n,\sigma}(t) - F_\sigma(t) \ge \frac{1}{\sqrt{18n}} \exp\left(-\frac{h^2}{4K^2}\right).$$

We first notice that for $0 < t < h$,

$$\tilde{F}_{n,\sigma}(t) - F_\sigma(t) = (\hat{p}_h - p_h)(\Phi_\sigma(t-h) - \Phi_\sigma(t)).$$

Letting $U_i = \mathbf{1}_{X_k = h}$, according to Berry-Esseen Theorem [Ber41, Ess56, Dur19], for $V \sim \mathcal{N}(0,1)$, we have

$$\sup_x \left| \mathbf{P}\left(\frac{1}{\sqrt{n\mathrm{Var}[U_1]}} \sum_{l=1}^n [U_l - \mathbb{E}U_1] \le -x\right) - \mathbf{P}(V \le -x) \right| \le \frac{\mathbb{E}|U_1 - \mathbb{E}[U_1]|^3}{2\sqrt{n}\sqrt{\mathrm{Var}[U_1]}^3}.$$

When $p_h < 1/2$, we have

$$\mathbb{E}[U_1] = p_h,$$
$$\mathrm{Var}[U_1] = p_h(1-p_h) \ge \frac{1}{2} p_h,$$
$$\mathbb{E}|U_1 - \mathbb{E}[U_1]|^3 \le \mathbb{E}|U_1|^3 = \mathbb{E}[U_1] = p_h.$$

We choose $x = 1$, and noticing that $P(V > 1) \ge \frac{1}{8}$ we obtain

$$\mathbf{P}\left(\hat{p}_h - p_h \le -\sqrt{\frac{p_h}{2n}}\right) = \mathbf{P}\left(\frac{1}{n}\sum_{l=1}^n U_l - \mathbb{E}[U_1] \le -\sqrt{\frac{p}{2n}}\right) \ge \frac{1}{8} - \frac{\mathbb{E}|U_1 - \mathbb{E}[U_1]|^3}{2\sqrt{n}\sqrt{\mathrm{Var}[U_1]}^3} \ge \frac{1}{8} - \frac{1}{\sqrt{2np_h}}.$$

This indicates that

$$\hat{p}_h - p_h \le -\frac{1}{\sqrt{2n}} \exp\left(-\frac{h^2}{4K^2}\right)$$

holds with probability at least $\frac{1}{8} - \frac{1}{\sqrt{2np_h}}$. Then due to the fact that when $0 < t < h-2$ and $h > \sigma$, $\Phi_\sigma(t-h) - \Phi_\sigma(t) \le \Phi_\sigma(0) - \Phi_\sigma(h) \le -\frac{1}{3}$, we have with probability at least $\frac{1}{8} - \frac{1}{\sqrt{2np_h}}$,

$$\tilde{F}_{n,\sigma}(t) - F_\sigma(t) \ge \frac{1}{\sqrt{18n}} \exp\left(-\frac{h^2}{4K^2}\right).$$

Therefore when $np_h \geq 128$ we will have the above inequality holds with probability at least $\frac{1}{16}$.

Combine these two results above, we know that whenever $h > \max\{C_h, \bar{h}, \sigma\}$ and n satisfies that

$$np_h \geq 128, \qquad \frac{1}{\sqrt{18n}} \exp\left(-\frac{h^2}{4K^2}\right) = \frac{4}{\sqrt{2\pi}\sigma} \exp\left(-(1-\delta)\zeta h^2\right), \tag{32}$$

we will have for $t = \frac{h}{2} + \frac{\sigma^2 h}{2K^2}$,

$$\tilde{F}_{n,\sigma}(t) - F_\sigma(t) \geq \frac{1}{\sqrt{18n}} \exp\left(-\frac{h^2}{4K^2}\right) \geq \frac{4}{\sqrt{2\pi}\sigma} \exp\left(-(1-\delta)\zeta h^2\right) \geq \mathbf{P}(X \in [t, t+2])$$

and hence $\tilde{F}_{n,\sigma}(t) \geq F_\sigma(t+2)$ holds with probability at least $\frac{1}{16}$. Hence Lemma 2 indicates that with probability at least $\frac{1}{16}$ we have

$$W_2(\mathbb{P}_h * \mathcal{N}(0,\sigma^2), \mathbb{P}_{h,n_h} * \mathcal{N}(0,\sigma^2))^2 \geq \mathbf{P}(X \in [t+1, t+2]) \geq \frac{1}{2\sqrt{2\pi}\sigma} \exp\left(-(1+\delta)\zeta h^2\right)$$

and hence

$$\mathbb{E}\left[W_2(\mathbb{P}_h * \mathcal{N}(0,\sigma^2), \mathbb{P}_{h,n_h} * \mathcal{N}(0,\sigma^2))\right] \geq \frac{1}{16\sqrt{2\sqrt{2\pi}\sigma}} \exp\left(-\frac{(1+\delta)\zeta h^2}{2}\right) = C_1 n^{-\alpha-\epsilon}, \tag{33}$$

where $C_1 = C_1(K, \sigma, \delta)$ is a positive constant. Here we use the fact that

$$\zeta = \frac{\left(\frac{1}{2} + \frac{\sigma^2}{2K^2}\right)^2}{2\sigma^2} \geq \frac{(\sigma/K)^2}{2\sigma^2} = \frac{1}{2K^2}$$

the second equation in (32) indicates that

$$h = \sqrt{\left((1-\delta)\zeta - \frac{1}{4K^2}\right)^{-1} \log \frac{12\sqrt{n}}{\sqrt{\pi}\sigma}} \tag{34}$$

and when $\delta < \frac{1}{2}$ and $n \geq \pi\sigma^2/144$ this is well-defined. Bringing in this formula of n into $\exp\left(-\frac{(1+\delta)\zeta h^2}{2}\right)$ will result in the RHS in (33), after noticing the definition of δ in (25).

Finally we show that for sufficiently large n, there always exists an h such that both (32) and also $h > \max\{C_h, \bar{h}, \sigma\}$ holds. For $n \geq \sqrt{\pi}\sigma/12$, we choose h in (34) and the second equation in (32) holds, and also when $\delta < \frac{1}{2}$ there exists n_0 such that for any $n > n_0$ we have $h > \max\{C_h, \bar{h}, \sigma\}$. With this choice of h, we further have

$$np_h = n \exp\left(-\frac{h^2}{2K^2}\right) \asymp n^{1-(4K^2(1-\delta)\zeta-1)^{-1}},$$

and since δ satisfies that $\delta < 1 - \frac{1}{2\zeta K^2}$ we know that

$$1 - (4K^2(1-\delta)\zeta - 1)^{-1} > 0.$$

Hence there exists a threshold n_0 such that for any $n > n_1$, we will have

$$np_h \geq 128.$$

Therefore, when $n > \max\{n_0, n_1\}$, with the choice of h in (34), we will have both (32) and also $h > \max\{C_h, \bar{h}, \sigma\}$ holds.

Therefore, for any $\epsilon > 0$, we have

$$\mathbb{E}\left[W_2(\mathbb{P}_h * \mathcal{N}(0,\sigma^2), \mathbb{P}_{h,n} * \mathcal{N}(0,\sigma^2))\right] = \Omega\left(n^{-\alpha-\epsilon}\right).$$

This completes the proof of Proposition 3. $\square$

4.2 Main Proof of the Lower Bound Part

The proof idea is similar to the above proof of Proposition 3. We summarize the properties of $\mathbb{P}_h$ for all $h > 0$ into one K-subgaussian distribution, such that this subgaussian distribution is a hard example for smoothed empirical W_2 convergence.

We construct the following discrete distribution

$$\mathbb{P} = \sum_{k=0}^{\infty} p_k \delta_{r_k}, \quad p_k \geq 0, \quad \sum_{k=1}^{\infty} p_k = 1, \tag{35}$$

where we choose $r_k = c_1 c_2 c_3 \cdots c_{k-1}$ for $k \geq 1$ for some positive constant $3 \leq c_1 \leq c_2 \leq c_3 \leq \cdots$ to be determined later, and

$$p_k = \frac{C}{\sqrt{2\pi} K} \exp\left(-\frac{r_k^2}{2K^2}\right), \quad k \geq 1, \quad p_0 = 1 - \sum_{k=1}^{\infty} p_k \tag{36}$$

where C is a small enough constant such that $\sum_{k=1}^{\infty} p_k \leq 1$. Then similar to the proof in Appendix A, we can prove that $\mathbb{P}$ is a K-subgaussian distribution.

We let $\kappa = \frac{\sigma^2}{K^2} \in (0,1)$, and

$$t_k = \frac{1}{2}(c_k + 1)(1+\kappa) \geq \frac{1}{2}(c_k+1) \geq 2.$$

First we will provide two propositions, which upper and lower bound the probability of $\mathbb{P} * \mathcal{N}(0,\sigma^2)$ near $t_k r_k$.

Proposition 4 (Probability Lower Bound)**.** *There exists some positive constant C_l only depending on σ and K such that*

$$\mathbf{P}(X \in [t_k r_k + 1, t_k r_k + 2]) \geq C_l \exp\left(-\left(t_k^2 - \kappa c_k - c_k\right) \cdot \frac{(r_k+2)^2}{2\sigma^2}\right), \quad X \sim \mathbb{P} * \mathcal{N}(0,\sigma^2).$$

Proof. We let $X = Y + Z$, where $Y \in \mathbb{P}, Z \sim \mathcal{N}(0,\sigma^2)$ are independent. Then we have

$$\begin{aligned}
\mathbf{P}(X \in [t_k r_k + 1, t_k r_k + 2]) &\geq \mathbf{P}(Y = r_k, Z \in [(t_k-1)r_k + 1, (t_k-1)r_k+2]) \\
&\geq p_k \cdot \mathbf{P}(Z \in [(t_k-1)r_k + 1, (t_k-1)r_k+2]) \\
&= \frac{1}{\sqrt{2\pi}\sigma} p_k \exp\left(-\frac{((t_k-1)r_k+2)^2}{2\sigma^2}\right) \\
&= \frac{C}{2\pi\sigma K} \exp\left(-\frac{r_k^2}{2K^2} - \frac{(t_k-1)^2(r_k+2)^2}{2\sigma^2}\right) \\
&\geq \frac{C}{2\pi\sigma K} \exp\left(-\left(\kappa + (t_k-1)^2\right) \cdot \frac{(r_k+2)^2}{2\sigma^2}\right) \\
&= \frac{C}{2\pi\sigma K} \exp\left(-\left(t_k^2 - \kappa c_k - c_k\right) \cdot \frac{(r_k+2)^2}{2\sigma^2}\right) \\
&\geq \frac{1}{2\pi\sigma K} \exp\left(-\left(t_k^2 - \kappa c_k - c_k\right) \cdot \frac{(r_k+2)^2}{2\sigma^2}\right),
\end{aligned}$$

where we use the fact that $C \geq 1$. Therefore, if we choose $C_l = \frac{1}{2\pi\sigma K}$, we have the desired lower bound in this proposition. ∎

Proposition 5 (Probability Upper Bound)**.** *When $c_k \geq \max\left\{\sqrt{\frac{2}{\kappa}}, \frac{\kappa+3}{1-\kappa}\right\}$, there exists some constant C_u only depending on K and σ such that*

$$\mathbf{P}(X \in [t r_k, t r_k + 2]) \leq C_u \exp\left(-(t_k^2 - c_k\kappa - c_k) \cdot \frac{(r_k-2)^2}{2\sigma^2}\right), \quad X \sim \mathbb{P} * \mathcal{N}(0,\sigma^2).$$

Proof. We let $X = Y + Z$, where $Y \in \mathbb{P}, Z \sim \mathcal{N}(0, \sigma^2)$ are independent. And we notice that

$$\begin{aligned}
\mathbf{P}(X \in [t_k r_k, t_k r_k + 2]) &= \sum_{j=0}^{\infty} \mathbf{P}(Y = r_j, Z \in [t_k r_k - r_j, t_k r_k - r_j + 2]) \\
&= \sum_{j=0}^{\infty} p_j \cdot \mathbf{P}(Z \in [t_k r_k - r_j, t_k r_k - r_j + 2]) \\
&\leq \frac{1}{\sqrt{2\pi}\sigma} \sum_{j=0}^{\infty} 2p_j \max\left\{\exp\left(-\frac{(t_k r_k - r_j)^2}{2\sigma^2}\right), \exp\left(-\frac{(t_k r_k - r_j + 2)^2}{2\sigma^2}\right)\right\} \\
&\leq \frac{2}{\sqrt{2\pi}\sigma} \sum_{j=0}^{k} p_j \exp\left(-\frac{(t_k r_k - r_j)^2}{2\sigma^2}\right) + \frac{2}{\sqrt{2\pi}\sigma} \sum_{j=k+1}^{\infty} p_j \exp\left(-\frac{(t_k r_k - r_j + 2)^2}{2\sigma^2}\right) \\
&\leq \sum_{j=1}^{k-1} \frac{2C}{\pi\sigma K} \exp\left(-\frac{r_j^2}{2K^2} - \frac{(t_k r_k - r_j)^2}{2\sigma^2}\right) + \frac{2C}{\pi\sigma K} \exp\left(-\frac{r_k^2}{2K^2} - \frac{(t_k r_k - r_k)^2}{2\sigma^2}\right) \\
&\quad + \sum_{j=k+1}^{\infty} \frac{2C}{\pi\sigma K} \exp\left(-\frac{r_j^2}{2K^2} - \frac{(t_k r_k - r_j + 2)^2}{2\sigma^2}\right) + \frac{2p_0}{\sqrt{2\pi}\sigma} \exp\left(-\frac{t_k^2 r_k^2}{2\sigma^2}\right).
\end{aligned}$$

Then we upper bound these three terms in the sum separately:

1. For $j = 0$, since $t_k^2 \geq t_k^2 - c_k\kappa - c_k$, with the choice $C_1 = \frac{2}{\sqrt{2\pi}\sigma}$, we have

$$\frac{2p_0}{\sqrt{2\pi}\sigma} \exp\left(-\frac{t_k r_k^2}{2\sigma^2}\right) \leq \frac{2}{\sqrt{2\pi}\sigma} \exp\left(-\frac{t_k r_k^2}{2\sigma^2}\right) \leq C_1 \exp\left(-\frac{(t_k^2 - \kappa c_k - c_k)(r_k - 2)^2}{2\sigma^2}\right).$$

2. For $j > k$, we have $r_j^2 \geq r_{k+1}^2 + j - (k+1)$. After noticing that $c_k \geq \frac{\kappa+3}{1-\kappa}$ and hence $c_k - t_k \geq 1$, we have

$$\begin{aligned}
\frac{(t_k r_k - r_j + 2)^2}{2\sigma^2} &\geq \frac{(r_{k+1} - t_k r_k - 2)^2}{2\sigma^2} = \frac{((c_k - t_k)r_k - 2)^2}{2\sigma^2} \\
&\geq \frac{(c_k - t_k)^2 (r_k - 2)^2}{2\sigma^2} = (t_k^2 - c_k\kappa - c_k - c_k^2\kappa) \cdot \frac{(r_k - 2)^2}{2\sigma^2}.
\end{aligned}$$

Therefore, choosing constant

$$C_2 = \sum_{j=0}^{\infty} \frac{2\sqrt{2K^2\pi} \exp\left(1/2K^2\right)}{\pi\sigma K} \exp\left(-\frac{j}{2K^2}\right) < \infty$$

and noticing that $C \le \sqrt{2K^2\pi}\exp\left(1/2K^2\right)$, we obtain:

$$
\begin{aligned}
&\sum_{j=k+1}^{\infty} \frac{2C}{\pi\sigma K}\exp\left(-\frac{r_j^2}{2K^2}-\frac{(t_jr_k-r_j+2)^2}{2\sigma^2}\right)\\
&\le \sum_{j=k+1}^{\infty} \frac{2C}{\pi\sigma K}\exp\left(-\frac{j-(k+1)+r_{k+1}^2}{2K^2}-\frac{(t_k^2-c_k\kappa-c_k-c_k^2\kappa)(r_k-2)^2}{2\sigma^2}\right)\\
&= \sum_{j=k+1}^{\infty} \frac{2C}{\pi\sigma K}\exp\left(-\frac{j-(k+1)}{2K^2}-\frac{\kappa c_k^2r_k^2}{2\sigma^2}-\frac{(t_k^2-c_k\kappa-c_k-c_k^2\kappa)(r_k-2)^2}{2\sigma^2}\right)\\
&\le \left(\sum_{j=k+1}^{\infty} \frac{2C}{\pi\sigma K}\exp\left(-\frac{j-(k+1)}{2K^2}\right)\right)\cdot\exp\left(-(t_k^2-c_k\kappa-c_k)\cdot\frac{(r_k-2)^2}{2\sigma^2}\right)\\
&\le \left(\sum_{j=k+1}^{\infty} \frac{2\sqrt{2K^2\pi}\exp\left(1/2K^2\right)}{\pi\sigma K}\exp\left(-\frac{j-(k+1)}{2K^2}\right)\right)\cdot\exp\left(-(t_k^2-c_k\kappa-c_k)\cdot\frac{(r_k-2)^2}{2\sigma^2}\right)\\
&= C_2\exp\left(-(t_k^2-c_k\kappa-c_k)\cdot\frac{(r_k-2)^2}{2\sigma^2}\right).
\end{aligned}
$$

3. For $j<k$, since $c_{k-1}\ge 3$, we first have

$$
\left(t_k-\frac{1}{c_{k-1}}\right)^2 \ge t_k^2-\frac{2t_k}{c_{k-1}} = t_k^2-\frac{(1+\kappa)(1+c_k)}{c_{k-1}} \ge t_k^2-\kappa c_k-c_k,
$$

   where in the last inequality we use the fact that $c_k\ge 1$ hence $\frac{1+c_k}{c_{k-1}}\le\frac{1+c_k}{2}\le c_k$. Therefore, noticing that $r_j\le\frac{r_k}{c_{k-1}}$, we obtain

$$
(t_kr_k-r_j)^2 \ge \left(t_k-\frac{1}{c_{k-1}}\right)^2 r_k^2 \ge (t_k^2-\kappa c_k-c_k)r_k^2.
$$

   Therefore, choosing constant $C_3=\sum_{j=1}^{\infty}\frac{2\sqrt{2K^2\pi}\exp\left(1/2K^2\right)}{\pi\sigma K}\exp\left(-\frac{j}{2K^2}\right)<\infty$, we will obtain:

$$
\begin{aligned}
&\sum_{j=1}^{k-1} \frac{2C}{\pi\sigma K}\exp\left(-\frac{r_j^2}{2K^2}-\frac{(t_kr_k-r_j)^2}{2\sigma^2}\right) \le \sum_{j=1}^{k-1} \frac{2C}{\pi\sigma K}\exp\left(-\frac{j}{2K^2}-\frac{(t_k^2-\kappa c_k-c_k)r_k^2}{2\sigma^2}\right)\\
&= \left(\sum_{j=1}^{k-1} \frac{2C}{\pi\sigma K}\exp\left(-\frac{j}{2K^2}\right)\right)\exp\left(-\frac{(t_k^2-\kappa c_k-c_k)r_k^2}{2\sigma^2}\right) \le C_3\exp\left(-\frac{(t_k^2-\kappa c_k-c_k)r_k^2}{2\sigma^2}\right)\\
&\le C_3\exp\left(-\frac{(t_k^2-\kappa c_k-c_k)(r_k-2)^2}{2\sigma^2}\right).
\end{aligned}
$$

4. For $j=k$, choosing $C_4=\frac{2\sqrt{2K^2\pi}\exp\left(1/2K^2\right)}{\pi\sigma K}$ and noticing $t_k=\frac{1}{2}(1+c_k)(1+\kappa)$, we will obtain:

$$
\begin{aligned}
\frac{2C}{\pi\sigma K}\exp\left(-\frac{r_k^2}{2K^2}-\frac{(t_kr_k-r_k)^2}{2\sigma^2}\right) &\le C_4\exp\left(-\left(t_k^2-\kappa c_k-c_k\right)\cdot\frac{r_k^2}{2\sigma^2}\right)\\
&\le C_4\exp\left(-\left(t_k^2-\kappa c_k-c_k\right)\cdot\frac{(r_k-2)^2}{2\sigma^2}\right).
\end{aligned}
$$

Therefore, choosing $C_u = C_1 + C_2 + C_3 + C_4$, we obtain:

$$\mathbf{P}(X \in [t_k r_k, t_k r_k + 2]) \leq C_u \exp\left(-(t_k^2 - c_k\kappa - c_k) \cdot \frac{(r_k - 2)^2}{2\sigma^2}\right).$$

□

We next present the following proposition, indicating that with positive probability the difference of CDFs of $\mathbb{P} * \mathcal{N}(0, \sigma^2)$ and $\mathbb{P}_n * \mathcal{N}(0, \sigma^2)$ is larger than $\frac{1}{2}\sqrt{\frac{p_{k+1}}{n}}$, which we will show is, in turn, larger than $\mathbb{P} * \mathcal{N}(0, \sigma^2)([t_k r_k, t_k r_k + 2])$ under some assumptions.

Proposition 6. *Suppose $c_k \geq \frac{\kappa+3}{1-\kappa}$ for every k. We use F_σ and $\tilde{F}_{n,\sigma}$ to denote the CDF of $P * \mathcal{N}(0, \sigma^2)$ and $P_n * \mathcal{N}(0, \sigma^2)$ respectively. Then there exists $k_0 = k_0(\sigma, K, C) > 0$ such that $\forall k \geq k_0$ and n with $np_{k+1} \geq 32768$, with probability at least $\frac{1}{64}$ we have*

$$\tilde{F}_{n,\sigma}(t_k r_k) - F_\sigma(t_k r_k) \geq \frac{1}{2}\sqrt{\frac{p_{k+1}}{n}}.$$

Proof. First we can write

$$F_\sigma(t_k r_k) = \sum_{j=0}^{\infty} p_j \Phi_\sigma(t_k r_k - r_j),$$
$$\tilde{F}_{n,\sigma}(t_k r_k) = \sum_{j=0}^{\infty} \hat{p}_j \Phi_\sigma(t_k r_k - r_j),$$

where Φ_σ is CDF of $\mathcal{N}(0, \sigma^2)$, and $\hat{p}_j$ is the empirical estimation of p_j with these n samples. Then we have

$$\begin{aligned}
\tilde{F}_{n,\sigma}(t_k r_k) - F_\sigma(t_k r_k) &= \sum_{j=0}^{\infty} (\hat{p}_j - p_j)\Phi_\sigma(t_k r_k - r_j) \\
&= \sum_{j=0}^{k} (\hat{p}_j - p_j)(1 - (1 - \Phi_\sigma(t_k r_k - r_j))) + \sum_{j=k+1}^{\infty} (\hat{p}_j - p_j)\Phi_\sigma(t_k r_k - r_j) \\
&\geq \sum_{j=0}^{k} \hat{p}_j - \sum_{j=0}^{k} p_j - \sum_{j=0}^{k} |\hat{p}_j - p_j|(1 - \Phi_\sigma(t_k r_k - r_j)) - \sum_{j=k+1}^{\infty} |\hat{p}_j - p_j|\Phi_\sigma(t_k r_k - r_j)
\end{aligned}$$

From assumption $c_k \geq \frac{\kappa+3}{1-\kappa}$ we know that $c_k \geq t_k + 1$. Hence for any $j \geq k+1$ we have $|t_k r_k - r_j| \geq |(c_k - t_k) r_k| \geq r_k \geq 1$ and for any $j \leq k$ we have $|t_k r_k - r_j| \geq (t_k - 1) r_j \geq r_j \geq 1$. According to the upper bound of Gaussian tail function (Proposition 2.1.2 in [Ver18]), we have

$$\begin{aligned}
1 - \Phi_\sigma(t_k r_k - r_j) &\leq \frac{1}{\sqrt{2\pi}} \cdot \frac{\sigma}{|t_k r_k - r_j|} \exp\left(-\frac{(t_k r_k - r_j)^2}{2\sigma^2}\right) \leq \sigma \exp\left(-\frac{(t_k r_k - r_j)^2}{2\sigma^2}\right), \quad \text{if } t_k r_k - r_j > 0, \\
\Phi_\sigma(t_k r_k - r_j) &\leq \frac{1}{\sqrt{2\pi}} \cdot \frac{\sigma}{|t_k r_k - r_j|} \exp\left(-\frac{(t_k r_k - r_j)^2}{2\sigma^2}\right) \leq \sigma \exp\left(-\frac{(t_k r_k - r_j)^2}{2\sigma^2}\right), \quad \text{if } t_k r_k - r_j < 0.
\end{aligned}$$

In the next, given that $np_{k+1} \geq 32768$, we will provide both a lower bound to $\sum_{j=0}^{k} \hat{p}_j - \sum_{j=0}^{k} p_j$ and also an upper bound to $|\hat{p}_{k+1} - p_{k+1}|$. As for $\sum_{j=0}^{k} \hat{p}_j - \sum_{j=0}^{k} p_j$, we can write it as

$$\sum_{j=0}^{k} \hat{p}_j - \sum_{j=0}^{k} p_j = \frac{1}{n}\left(\sum_{l=1}^{n} U_l\right) - \mathbb{E}[U_1],$$

where $U_l \sim \mathrm{Bern}(\sum_{j=k+1}^{\infty} p_j)$ are *i.i.d.* Bernoulli random variables. According to the Berry-Esseen Theorem [Dur19] we have

$$\left|\mathbf{P}\left(\frac{1}{\sqrt{n\mathrm{Var}[U_1]}}\sum_{l=1}^{n}[U_l - \mathbb{E}U_1] \geq 1\right) - \mathbf{P}(V \geq 1)\right| \leq \frac{\mathbb{E}|U_1 - \mathbb{E}[U_1]|^3}{2\sqrt{n}\sqrt{\mathrm{Var}[U_1]}^3}$$

where $V \sim \mathcal{N}(0,1)$. It is easy to check that $\sum_{j=k+1}^{\infty} p_j \leq 2p_{j+1} < 1/2$ for $k \geq 2$. Hence we have

$$\mathrm{Var}[U_1] = \left(\sum_{j=k+1}^{\infty} p_j\right)\left(1 - \sum_{j=k+1}^{\infty} p_j\right) \geq \frac{1}{2}\left(\sum_{j=k+1}^{\infty} p_j\right) \geq \frac{1}{2}p_{k+1}$$

$$\mathbb{E}|U_1 - \mathbb{E}[U_1]|^3 \leq \mathbb{E}|U_1|^3 = \mathbb{E}[U_1] = \sum_{j=k+1}^{\infty} p_j \leq 2p_{k+1}.$$

Noticing that for standard Gaussian random variable $V \sim \mathcal{N}(0,1)$ we have $P(V > 1) \geq 1/8$, we obtain that

$$\begin{aligned}
\mathbf{P}\left(\sum_{j=0}^{k}\hat{p}_j - \sum_{j=0}^{k} p_j \geq \sqrt{\frac{p_{k+1}}{2n}}\right) &= \mathbf{P}\left(\frac{1}{n}\sum_{l=1}^{n} U_l - \mathbb{E}[U_1] \geq \sqrt{\frac{p_{k+1}}{2n}}\right) \\
&\geq \mathbf{P}\left(\frac{1}{\sqrt{n\mathrm{Var}[U_1]}}\sum_{l=1}^{n} U_l - \mathbb{E}[U_1] \geq 1\right) \\
&\geq \frac{1}{8} - \frac{\mathbb{E}|U_1 - \mathbb{E}[U_1]|^3}{2\sqrt{n}\sqrt{\mathrm{Var}[U_1]}^3} \geq \frac{1}{8} - \frac{2\sqrt{2}}{\sqrt{np_{k+1}}} \geq \frac{1}{16},
\end{aligned}$$

where we use the fact that $np_{k+1} \geq 32768$. As for $|\hat{p}_{k+1} - p_{k+1}|$, if we let $U_l' \sim \mathrm{Bern}(p_{k+1}), i.i.d$, again according to Berry-Esseen [Dur19] Theorem we obtain that

$$\begin{aligned}
\mathbf{P}\left(|\hat{p}_{k+1} - p_{k+1}| \geq 8\sqrt{\frac{p_{k+1}}{n}}\right) &= \mathbf{P}\left(\frac{1}{n}\sum_{l=1}^{n} U_l' - \mathbb{E}[U_1'] \geq 8\sqrt{\frac{p_{k+1}}{n}}\right) \\
&\leq \mathbf{P}\left(\frac{1}{\sqrt{n\mathrm{Var}[U_1']}}\sum_{l=1}^{n} U_l' - \mathbb{E}[U_1'] \geq 8\right) \\
&\leq \frac{1}{128} + \frac{\mathbb{E}|U_1' - \mathbb{E}[U_1']|^3}{2\sqrt{n}\sqrt{\mathrm{Var}[U_1']}^3} \leq \frac{\sqrt{2}\mathbb{E}|U_1'|^3}{\sqrt{n}\sqrt{p_{k+1}}^3} \leq \frac{1}{128} + \frac{\sqrt{2}}{\sqrt{np_{k+1}}} \leq \frac{1}{64}
\end{aligned}$$

after noticing that $\mathrm{Var}[U_1] = p_{k+1}(1 - p_{k+1}) \leq p_{k+1}$ and also $P(V > 8) \leq 1/128$ for $V \sim \mathcal{N}(0,1)$, and the last inequality follows from $np_{k+1} \geq 32768$.

We further notice that

$$\mathbb{E}\left[\max_{j\geq 0}|\hat{p}_j - p_j|^2\right] \leq \mathbb{E}\left[\sum_{j=0}^{\infty}|\hat{p}_j - p_j|^2\right] = \sum_{j=0}^{\infty}\mathrm{Var}(\hat{p}_j) = \sum_{j=0}^{\infty}\frac{p_j(1-p_j)}{n} \leq \frac{1}{n}.$$

Hence by the Markov inequality we obtained that

$$\mathbf{P}\left(\max_{j\geq 0}|\hat{p}_j - p_j| \leq \frac{4}{\sqrt{n}}\right) \geq \frac{15}{16}. \tag{37}$$

Therefore, if $n \geq 32768/p_{k+1}$, according to (37), with probability at least $\frac{1}{64}$ we have

$$\tilde{F}_{n,\sigma}(t_k r_k) - F_\sigma(t_k r_k) \geq \sqrt{\frac{p_{k+1}}{2n}} - \frac{4\sigma}{\sqrt{n}} \sum_{j=0}^{k} \exp\left(-\frac{(t_k r_k - r_j)^2}{2\sigma^2}\right) - \frac{4\sigma}{\sqrt{n}} \sum_{j=k+2}^{\infty} \exp\left(-\frac{(t_k r_k - r_j)^2}{2\sigma^2}\right) \tag{38}$$

$$- 8\sigma\sqrt{\frac{p_{k+1}}{n}} \exp\left(-\frac{(t_k r_k - r_{k+1})^2}{2\sigma^2}\right). \tag{39}$$

Additionally, we have

$$\sum_{j=0}^{k} \exp\left(-\frac{(t_k r_k - r_j)^2}{2\sigma^2}\right) \leq k \exp\left(-\frac{(t_k - 1)^2 r_k^2}{2\sigma^2}\right).$$

And for any $j \geq k+2$, we have $r_j - t_k r_k \geq j - (k+2) + r_{k+2} - t_k r_k \geq j - (k+2) + (t_k - 1)t_k$, which indicates that

$$\sum_{j=k+2}^{\infty} \exp\left(-\frac{(t_k r_k - r_j)^2}{2\sigma^2}\right) \leq \left(\sum_{j=k+2}^{\infty} \exp\left(-\frac{j-(k+2)}{2\sigma^2}\right)\right) \cdot \exp\left(-\frac{(t_k - 1)^2 r_k^2}{2\sigma^2}\right) \leq C_j \exp\left(-\frac{(t_k - 1)^2 r_k^2}{2\sigma^2}\right),$$

where C_j is a constant only depending on σ. We also notice that $\frac{(t_k r_k - r_{k+1})^2}{2\sigma^2} \geq \frac{r_k^2}{2\sigma^2}$ using the fact that $c_k \geq t_k + 1$, and that

$$\exp\left(-\frac{(t_k - 1)^2 r_k^2}{2\sigma^2}\right) \leq \exp\left(-\frac{c_k^2 r_k^2}{4K^2} - \frac{c_k^2 r_k^2 \kappa^2}{8\sigma^2}\right) = \sqrt{\frac{\sqrt{2\pi} K p_{k+1}}{C}} \cdot \exp\left(-\frac{c_k^2 \kappa^2 r_k^2}{8\sigma^2}\right)$$

using the fact that

$$2c_k^2\kappa + c_k^2\kappa^2 \leq c_k^2\kappa^2 + c_k^2 + \kappa^2 + 1 - 2c_k - 2\kappa + 2c_k^2\kappa = (2t_k - 2)^2.$$

Hence we have

$$\frac{4\sigma}{\sqrt{n}} \sum_{j=0}^{k} \exp\left(-\frac{(t_k r_k - r_j)^2}{2\sigma^2}\right) + \frac{4\sigma}{\sqrt{n}} \sum_{j=k+2}^{\infty} \exp\left(-\frac{(t_k r_k - r_j)^2}{2\sigma^2}\right) + 8\sigma\sqrt{\frac{p_{k+1}}{n}} \exp\left(-\frac{(t_k r_k - r_{k+1})^2}{2\sigma^2}\right)$$
$$\leq 4\sqrt{\frac{p_{k+1}}{n}} \cdot \sigma \left(\frac{C_j\sqrt{\sqrt{2\pi}K} + k}{\sqrt{C}} \exp\left(-\frac{c_k^2\kappa^2 r_k^2}{8\sigma^2}\right) + \exp\left(-\frac{r_k^2}{2\sigma^2}\right)\right).$$

Since $r_k = c_1 c_2 \cdots c_{k-1} \geq 3^{k-1}$, there exists some constant k_0 only depending on K, σ, C such that for any $k \geq k_0$, we have

$$\sigma \left(\frac{C_j\sqrt{\sqrt{2\pi}K} + k}{\sqrt{C}} \exp\left(-\frac{c_k^2\kappa^2 r_k^2}{8\sigma^2}\right) + \exp\left(-\frac{r_k^2}{2\sigma^2}\right)\right) \leq \frac{1}{4\sqrt{2}} - \frac{1}{8}$$

Bringing this result to (38), we will obtain that for any $k \geq k_0$,

$$\tilde{F}_{n,\sigma}(t_k r_k) - F_\sigma(t_k r_k) \geq \frac{1}{2}\sqrt{\frac{p_{k+1}}{n}}$$

holds. This completes the proof of this proposition. ☐

With the above propositions, we are now ready to prove the lower bound part of Theorem 2.

Proof of Lower Bound Part of Theorem 2. We let $t_k = \frac{1}{2}(1+\kappa)(1+c_k)$ and

$$n_k = \left\lfloor \frac{1}{4C_u^2} \exp\left((t_k^2 - c_k\kappa - c_k) \cdot \frac{(r_k-2)^2}{\sigma^2} - \frac{c_k^2 r_k^2}{2K^2} \right) \right\rfloor, \tag{40}$$

Then there exists some constant k_0' only depending on k, σ and C such that for any $k \geq k_0'$, we would have

$$n_k p_{k+1} \geq 32768.$$

Hence according to Proposition 6 we would have when $k \geq \max\{k_0, k_0'\}$,

$$\tilde{F}_{n_k,\sigma}(t_k r_k) - F_\sigma(t_k r_k) \geq \frac{1}{2}\sqrt{\frac{p_{k+1}}{n_k}}$$

holds with probability at least $\frac{1}{64}$. Moreover, with our choice of n_k, it is easy to check that

$$\frac{1}{2}\sqrt{\frac{p_{k+1}}{n_k}} \geq C_u \exp\left(-(t_k^2 - c_k\kappa - c_k) \cdot \frac{(r_k-2)^2}{2\sigma^2} \right).$$

Hence according to Proposition 5, with probability at least $\frac{1}{64}$ we have for $X \sim \mathbb{P} * \mathcal{N}(0,\sigma^2)$,

$$\tilde{F}_{n_k,\sigma}(t_k r_k) - F_\sigma(t_k r_k) \geq C_u \exp\left(-(t_k^2 - c_k\kappa - c_k) \cdot \frac{(r_k-2)^2}{2\sigma^2} \right) \geq \mathbf{P}(X \in [t_k r_k, t_k r_k + 2]).$$

Therefore we have

$$\tilde{F}_{n_k,\sigma}(t_k r_k) \geq F_\sigma(t_k r_k + 2).$$

According to Lemma 2 and Proposition 4, this indicates that with probability at least $\frac{1}{64}$,

$$\begin{aligned} W_2(\mathbb{P} * \mathcal{N}(0,\sigma^2), \mathbb{P}_{n_k} * \mathcal{N}(0,\sigma^2)) &\geq \sqrt{\mathbf{P}(X \in [t_k r_k + 1, t_k r_k + 2])} \\ &\geq \sqrt{C_l \exp\left(-(t_k^2 - c_k\kappa - c_k) \cdot \frac{(r_k+2)^2}{2\sigma^2} \right)}, \end{aligned}$$

where $X \sim \mathbb{P}_{n_k} * \mathcal{N}(0,\sigma^2)$. Hence we obtain that

$$\mathbb{E}[W_2(\mathbb{P} * \mathcal{N}(0,\sigma^2), \mathbb{P}_{n_k} * \mathcal{N}(0,\sigma^2))] \geq \frac{\sqrt{C_l}}{64} \sqrt{\exp\left(-(t_k^2 - c_k\kappa - c_k) \cdot \frac{(r_k+2)^2}{2\sigma^2} \right)},$$

which indicates that there exists some constant C_5, C_6 only depending on C, K, σ such that

$$\frac{\mathbb{E}[W_2(\mathbb{P} * \mathcal{N}(0,\sigma^2), \mathbb{P}_{n_k} * \mathcal{N}(0,\sigma^2))]}{n_k^{-\frac{(t^2 - c\kappa - c)/(4\sigma^2)}{(t^2 - c\kappa - c)/\sigma^2 - c^2/(2K^2)}}} \geq C_5 \exp\left(-C_6 r_k\right) \geq n_k^{-\mathcal{O}\left(\frac{1}{\sqrt{\log n_k}}\right)}.$$

Next we remember that $t_k = \frac{1}{2}(1+\kappa)(1+c_k)$, therefore if choosing c large enough, we will have

$$t_k^2 - c_k\kappa - c_k = \frac{(1+\kappa)^2(1+c_k)^2}{4} - c_k(1+\kappa) = \frac{(1+\kappa)^2 c_k^2}{4} + \mathcal{O}(c_k),$$

which indicates that

$$\frac{(t^2 - c\kappa - c)/(4\sigma^2)}{(t^2 - c\kappa - c)/\sigma^2 - c^2/(2K^2)} = \frac{(1+\kappa)^2 c_k^2 + \mathcal{O}(c_k)}{4(1+\kappa)^2 c_k^2 + \mathcal{O}(c_k) - 8c_k^2\kappa} = \frac{(1+\kappa)^2}{4(1+\kappa)^2 - 8\kappa} + \mathcal{O}\left(\frac{1}{c_k}\right) = \alpha + \mathcal{O}\left(\frac{1}{c_k}\right).$$

Therefore, choosing $c_k = M^k$ with $M = \max\left\{\sqrt{\frac{2}{\kappa}}, \frac{\kappa+3}{1-\kappa}, 3\right\}$, then for every k we have $c_k \geq \max\left\{\sqrt{\frac{2}{\kappa}}, \frac{\kappa+3}{1-\kappa}, 3\right\}$, which indicates that this choice of c_k satisfies all previous assumptions on c_k. We further notice that $r_k = M^{\frac{k(k-1)}{2}}$, hence

$$\frac{n_k^{-\frac{(t^2-c\kappa-c)/(4\sigma^2)}{(t^2-c\kappa-c)/\sigma^2-c^2/(2K^2)}}}{n_k^{-\alpha}} \geq n_k^{-\mathcal{O}\left(\frac{1}{\sqrt{\log\log n_k}}\right)}.$$

Therefore, we obtain that

$$n^\alpha \mathbb{E}[W_2(\mathbb{P} * \mathcal{N}(0,\sigma^2), \mathbb{P}_{n_k} * \mathcal{N}(0,\sigma^2))] \leq n_k^{-\mathcal{O}\left(\frac{1}{\sqrt{\log\log n_k}}\right)}.$$

We let k goes to infinity, and obtain that

$$\limsup_{n\to\infty} n^{\alpha+\frac{1}{\sqrt[3]{\log\log n}}} \mathbb{E}[W_2(\mathbb{P} * \mathcal{N}(0,\sigma^2), \mathbb{P}_n * \mathcal{N}(0,\sigma^2))] \geq \limsup_{k\to\infty} n_k^{\frac{1}{\sqrt[3]{\log\log n_k}} - \mathcal{O}\left(\frac{1}{\sqrt{\log\log n_k}}\right)} > 0.$$

And the proof of the lower bound part of Theorem 2 is completed. □

5 Proof of the Upper Bound of Theorem 2

Without loss of generality, we consider the case $\sigma = 1$, as we can always reduce to this by rescaling. We start the proof from the following observation [Vil03, Theorem 2.18]: for two distributions $\mathbb{Q}_1, \mathbb{Q}_2$ on $\mathbb{R}$ with absolutely continuous CDFs $F_1(x), F_2(x)$ the optimal coupling for the 2-Wasserstein distance is given by $F_2^{-1}(F_1(x))$, implying an explicit formula:

$$W_2(\mathbb{Q}_1, \mathbb{Q}_2)^2 = \mathbb{E}[\left|F_2^{-1}(F_1(T_1)) - T_1\right|^2], T_1 \sim \mathbb{Q}_1,$$

where $F_2^{-1}(\cdot)$ is the inverse function of $F_2(\cdot)$. In our case we set $\mathbb{Q}_1 = \mathbb{P} * \mathcal{N}(0,1)$ and $\mathbb{Q}_2 = \mathbb{P}_n * \mathcal{N}(0,1)$, and denote their CDFs by F and F_n respectively. In the following, we use $\mathcal{N}$ to denote $\mathcal{N}(0,1)$. We also denote by $\rho(t)$ the pdf of F and by T the optimal transport map

$$T(t) \triangleq F_n^{-1}(F(t)).$$

Note that because of the randomness of F_n the map T is random as well. Our proof will proceed along the following reductions:

1. (Conditioning) Note that if E is any event with probability at least $1 - O(\frac{1}{n^2})$ then we have

$$\mathbb{E}[W_2^2(\mathbb{P} * \mathcal{N}, \mathbb{P}_n * \mathcal{N})] \leq \mathbb{E}[W_2^2(\mathbb{P} * \mathcal{N}, \mathbb{P}_n * \mathcal{N})|E] + O(\frac{1}{n}) \tag{41}$$

   This allows us to condition on a typical realization of the empirical measure $\mathbb{P}_n$.

2. (Truncation) We will show that with high probability

$$|T(t) - t| \lesssim |t| + \sqrt{\log n} \qquad \forall t \in \mathbb{R}. \tag{42}$$

   Conditioning on this event, then, allows us to restrict evaluation of W_2^2 to a $O(\log n)$ range of t:

$$W_2^2(\mathbb{P} * \mathcal{N}, \mathbb{P}_n * \mathcal{N}) = \mathbb{E}[|T(X+Z) - (X+Z)|^2 1\{|X+Z| \leq b\sqrt{\log n}] + O(1/n).$$

3. (Key bound) So far we are left to bound the integral

$$\int_{|t|\le b\sqrt{\log n}} \rho(t)|T(t)-t|^2 dt \tag{43}$$

and we only have the bound (42). The *key novel ingredient* is the following observation. The transport distance can be bounded by

$$|T(t)-t| \le \frac{|F(t)-F_n(t)|}{\rho(t)} \tag{44}$$

This bound can be explained by the fact that if $F(t) < F_n(t)$ then the distance we need to travel to the right of t so that $F(\cdot)$ raise to the value of $F_n(t)$ will be around $\frac{|F(t)-F_n(t)|}{\rho(t)}$. There are several caveats in a rigorous statement of (44) (see Prop. 8 for details), but the most important one is that it only holds provided the RHS of (44) is ≤ 1.

4. (Concentration) Next, we will show that with high probability

$$|F_n(t)-F(t)| \lesssim \frac{\log n}{\sqrt{n}}\sqrt{\min(F(t),1-F(t))\vee\frac{1}{n}}\,. \tag{45}$$

It turns out that we also have $\min(F(t),1-F(t)) \lesssim \rho(t)^{\frac{4K^2}{(1+K^2)^2}} n^{o(1)}$. Hence, we have a transport distance bound

$$|T(t)-t| \lesssim \frac{\log n}{\sqrt{n}}\rho(t)^{\frac{2K^2}{(1+K^2)^2}-1} n^{o(1)}\,,$$

provided the RHS is $\lesssim 1$, which is equivalent to say $\rho(t) > n^{-\alpha-o(1)}$ for $\alpha = \frac{(1+K^2)^2}{2(1+K^4)}$. Note that in this region the integral (43) becomes bounded by

$$n^{o(1)}\int_{|t|\le b\sqrt{\log n},\rho(t)>n^{-\alpha}} \frac{1}{n}\rho^{\frac{4K^2}{(1+K^2)^2}-1}(t) \le n^{-\alpha+o(1)}\,, \tag{46}$$

since $K>1$ and thus the power of ρ is negative.

5. (Final) The final step is to split the integral (43) into values of $\rho(t) < n^{-\alpha}$ (for which we use the bound (42) and $|t| \lesssim \sqrt{\log n}$) and $\rho(t) > n^{-\alpha}$ (for which we use (46)). This gives us the contributions

$$n^{-\alpha}O(\log n) + n^{o(1)}n^{-\alpha}$$

completing the proof.

If we can prove for any $\epsilon > 0$, there exists C_ϵ such that for any n and K-subgaussian distribution $\mathbb{P}$,

$$\mathbb{E}[W_2^2(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})] \le C_\epsilon n^{-2\alpha+\epsilon}, \tag{47}$$

then for every integer t and n we have

$$\mathbb{E}[W_2^2(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})] \le C_{1/(2t)} n^{-2\alpha+1/(2t)}.$$

WLOG we assume that $C_{1/(2t)} \ge C_{1/(2s)} \ge 1$ for every $t > s$. Therefore, when $n \ge C_{1/(2t)}^{2t}$ we have $\mathbb{E}[W_2^2(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})] \le n^{-2\alpha+1/t}$ for all K-subgaussian distribution $\mathbb{P}$. We let $\delta_n = 1/(2t)$ for those $n \in (C_{1/(2(t-1))}^{2(t-1)}, C_{1/(2t)}^{2t}]$, and for those $n \le C_{1/2}^2$, we choose $\delta_n = \log_2\left[\max_{2\le n\le C_{1/2}^2}\mathbb{E}[W_2^2(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})]\right]$, we will have

$$\mathbb{E}[W_2^2(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})] \le n^{-2\alpha+\delta_n}$$

with $\lim_{n\to\infty}\delta_n = 0$. Therefore, we only need to prove (47)

Proposition 7. *We denote the CDF, PDF of $\mathbb{P} * \mathcal{N}(0,1)$ as F, ρ, respectively, and let $X \sim \mathbb{P}$. Suppose there exist constants $C, K > 0$ such that for $\forall r \geq 0$,*

$$\mathbf{P}(|X| \geq r) \leq C \exp\left(-\frac{r^2}{2K^2}\right).$$

For $\beta = \frac{4K^2}{(1+K^2)^2}$ and any $0 < \epsilon < \beta$, $\exists M = M(K, C, \epsilon) \geq 1$ such that for any K-subgaussian distribution $\mathbb{P}$,

$$\begin{aligned} 1 - F(r) &\leq M\rho(r)^{\beta-\epsilon}, \quad \forall r \geq 0, \\ F(r) &\leq M\rho(r)^{\beta-\epsilon}, \quad \forall r < 0. \end{aligned}$$

Remark 4. *We notice that this result is tight when considering $\mathbb{P}_h = (1-p_h)\delta_0 + p_h\delta_h$ with $p_h = \exp\left(-\frac{h^2}{2K^2}\right)$ and $r = \frac{(K^2+1)h}{2K^2}$. Then we have $\rho(r) \asymp \exp\left(-\frac{(K^2+1)^2h^2}{8K^4}\right)$ and $1 - F(r) \asymp \exp\left(-\frac{h^2}{2K^2}\right)$. Hence the above inequalities are tight.*

First we present two lemmas:

Lemma 3. *Suppose Φ_1 to be the CDF of Gaussian distribution $\mathcal{N}(0,1)$, then we have*

$$\begin{aligned} 1 - \Phi_1(l) &\leq \exp\left(-\frac{l^2}{2}\right), \quad \forall l \geq 0 \\ \Phi_1(l) &\leq \exp\left(-\frac{l^2}{2}\right), \quad \forall l < 0 \end{aligned}$$

Proof. Since we have $\Phi_1(l) = 1 - \Phi_1(l)$ for any $l \geq 0$, we only need to prove the results for $l \geq 0$. According to the upper bound on the tail of Gaussian distributions [Ver18, Proposition 2.1.2], we have for $l \geq 1$,

$$1 - \Phi_1(l) \leq \frac{1}{l} \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{l^2}{2}\right) \leq \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{l^2}{2}\right) \leq \exp\left(-\frac{l^2}{2}\right).$$

For $0 \leq l \leq 1$, we have

$$1 - \Phi_1(l) \leq 1 - \frac{1}{2} = \frac{1}{2}, \quad \exp\left(-\frac{l^2}{2}\right) \geq \exp(-1/2) \geq \frac{1}{2},$$

which indicates that

$$1 - \Phi_1(l) \leq \frac{1}{2} \leq \exp\left(-\frac{l^2}{2}\right).$$

Hence for $\forall l \geq 0$,

$$1 - \Phi_1(l) \leq \exp\left(-\frac{l^2}{2}\right).$$

$\square$

Lemma 4. *Suppose $\mathbb{P}$ is a 1-dimensional K-subgaussian distribution (for some constant $C > 0$ we have $\mathbf{P}[X \geq r] \leq C \exp\left(-\frac{r^2}{2K^2}\right)$ for every $r \geq 0$, where $X \sim \mathbb{P}$), and $\rho(\cdot)$ is the PDF of $\mathbb{P} * \mathcal{N}$. For any $0 < \epsilon < \beta$ we have*

$$\rho(r) \geq C_\epsilon \mathbf{P}[X \geq r]^{\frac{1}{\beta-\epsilon}} \quad \forall r \geq 0$$

for some positive constant $C_\epsilon = C_\epsilon(K, C)$.

Proof. For $X \sim \mathbb{P}$, choosing $M = M(K, C) \triangleq K\sqrt{2\log(2C)} > 0$ then we have

$$\mathbf{P}[X \in [-M, M]] = 1 - \mathbf{P}[|X| > M] \leq 1 - C\exp\left(-\frac{M^2}{2K^2}\right) \geq \frac{1}{2}.$$

For $0 \leq r \leq M$, we have

$$\rho(r) \geq \int_{-M}^{M} \eta(x)\varphi_1(r-x)dx \geq \mathbf{P}[X \in [-M, M]] \cdot \min_{-M \leq x \leq M} \varphi(r-x) \geq \frac{1}{2}\varphi_1(2M),$$

where we use $\eta(\cdot)$ to denote the PDF of $\mathbb{P}$ (which can be a generalized function). Hence $\rho(r) \geq C_\epsilon \mathbf{P}[X \geq r]^{\frac{1}{\beta-\epsilon}}$ holds for all $0 \leq r \leq M$ if $C_\epsilon \leq \frac{1}{2}\varphi_1(2M)$.

Next we consider cases where $r \geq M$. We let $c_r = \log \frac{C}{\mathbf{P}[X \geq r]}$. If $c_r \geq \log \frac{1}{\rho(r)}$, then we have

$$\rho(r) \geq \frac{\mathbf{P}[X \geq r]}{C} \geq \frac{1}{C}\mathbf{P}[X \geq r]^{\frac{1}{\beta-\epsilon}},$$

where we use the fact that $\beta - \epsilon \leq \beta \leq 1$ and hence $\frac{1}{\beta-\epsilon} \geq 1$.

Next we consider cases where $c_r < \log \frac{1}{\rho(r)}$. We let $r_1 = \sqrt{2\log(2) + 2c_r}K$, then we have $\mathbf{P}[X \geq r] = Ce^{-c_r}$, and

$$\mathbf{P}[X \geq r_1] \leq C\exp\left(-\frac{r_1^2}{2K^2}\right) \leq \frac{C}{2}e^{-c_r}.$$

Hence $\mathbf{P}[r < X \leq r_1] \geq \frac{C}{2}e^{-c_r}$, which indicates that

$$\begin{aligned}\rho(r) &\geq \int_{-M}^{M} \eta(x)\varphi(r-x)dx + \int_{r}^{r_1} \eta(x)\varphi(r-x)dx \geq \frac{1}{2}\varphi(r+M) + \frac{C}{2}e^{-c_r}\varphi(r_1 - r) \\ &= \frac{1}{2\sqrt{2\pi}}\exp\left(-\frac{(r+M)^2}{2}\right) + \frac{C}{2\sqrt{2\pi}}\exp\left(-c_r - \frac{(r_1-r)^2}{2}\right).\end{aligned}$$

Next, we let $c_r = \frac{x^2r^2}{2K^2}$ with $x \geq 1$. We notice that when $x \geq \frac{2K^2}{1+K^2}$ we have $\frac{x^2}{\beta K^2} \geq 1$, hence $-\frac{r^2}{2} \geq -\frac{1}{\beta}c_r$; and when $1 \leq x \leq \frac{2K^2}{1+K^2}$, we have $\left(\frac{1}{\beta K^2} - \frac{1}{K^2} - 1\right)x^2 + 2x - 1 \geq 0$ since $-1 \leq \frac{1}{\beta K^2} - \frac{1}{K^2} - 1 = \frac{(K^2+1)(1-3K^2)}{4K^4} \leq 0$, and hence $-c_r - \frac{(r-\sqrt{2c_r}K)^2}{2} \geq -\frac{1}{\beta}c_r$. Therefore, we have

$$\max\left\{-\frac{r^2}{2}, -c_r - \frac{(r-\sqrt{2c_r}K)^2}{2}\right\} \geq -\frac{1}{\beta}c_r$$

We further notice that

$$\rho(r) = \int_{-\infty}^{\infty} \eta(x)\varphi_1(r-x)dx = \int_{-\infty}^{r/2} \eta(x)\varphi_1(r-x)dx + \int_{r/2}^{\infty} \eta(x)\varphi_1(r-x)dx \tag{48}$$

$$\leq \sup_{x \leq r/2} \varphi_1(r-x) + \mathbf{P}[X \geq r/2] \leq \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{r^2}{8}\right) + C\exp\left(-\frac{r^2}{8K^2}\right) \leq \left(\frac{1}{\sqrt{2\pi}} + C\right)\exp\left(-\frac{r^2}{8K^2}\right). \tag{49}$$

This indicates that $r \leq 2\sqrt{K\log\frac{\bar{C}}{\rho(r)}}$ with $\bar{C} = \frac{1}{\sqrt{2\pi}} + C$, and hence $\exp(rM) = \mathcal{O}(\rho(r)^{-\epsilon'})$ for $\forall \epsilon' > 0$. We further notice that $\exp\left(2\sqrt{\log(2)}K^2\sqrt{c_r}\right) = \mathcal{O}(\rho(r)^{-\epsilon'})$ for $\forall \epsilon' > 0$ since $c_r \leq \log\frac{1}{\rho(r)}$. Therefore, we obtain that

$$
\begin{aligned}
&\mathbf{P}[X \geq r]^{\frac{1}{\beta}} = C^{(1/\beta)} \exp\left(-\frac{1}{\beta} c_r\right) \\
&\leq C^{(1/\beta)} \max\left\{\exp\left(-\frac{r^2}{2}\right), \exp\left(-c_r - \frac{(r-\sqrt{2c_r}K)^2}{2}\right)\right\} \\
&\leq \frac{2\sqrt{2\pi}C^{(1/\beta)}}{1+C} \cdot \left(\frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{r^2}{2}\right) + \frac{C}{2\sqrt{2\pi}} \exp\left(-c_r - \frac{(r-\sqrt{2c_r}K)^2}{2}\right)\right) \\
&\leq \rho(r) \cdot \max\left\{\exp\left(Mr + \frac{M^2}{2}\right), \exp\left(\sqrt{2\log(2)}K(\sqrt{2c_r}K - r) + 2K^2 \log 2\right)\right\} \\
&\leq \rho(r) \cdot \max\left\{\exp\left(Mr + \frac{M^2}{2}\right), \exp\left(2\sqrt{\log(2)}K^2\sqrt{c_r} + 2K^2 \log 2\right)\right\} \\
&\leq \rho(r) \cdot \tilde{\mathcal{O}}(\rho^{-\epsilon'}).
\end{aligned}
$$

Choosing $\epsilon' = \frac{\epsilon}{\beta}$, we know that there exists some positive constant C_ϵ such that

$$
\rho(r) \geq C_\epsilon \mathbf{P}[X \geq r]^{\frac{1}{\beta-\epsilon}} \quad \forall r \geq 0.
$$

□

Proof of Proposition 7. We only prove this results for $r \geq 0$, as the proof of $r \leq 0$ is similar. First we can write

$$
1 - F(r) = \int_{-\infty}^{\infty} \eta(t)(1 - \Phi_1(r-t))dt, \tag{50}
$$

$$
\rho(r) = \int_{-\infty}^{\infty} \eta(t) \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(r-t)^2}{2}\right) dt. \tag{51}
$$

Noticing that $\mathbf{P}(|X| \geq r) \leq C \exp\left(-\frac{r^2}{2K^2}\right)$, If we choose

$$
\tilde{K} = K\sqrt{2(\log(2C))},
$$

we will obtain that

$$
\mathbf{P}(|X| \geq \tilde{K}) \leq C \exp(-\log(2C)) = \frac{1}{2}
$$

and hence $\mathbf{P}(|X| \leq \tilde{K}) \geq \frac{1}{2}$. In the following, we will discuss cases where $0 \leq r \leq \tilde{K}$ and $r > \tilde{K}$ separately.

If $0 \leq r \leq \tilde{K}$, then we have

$$
\begin{aligned}
\rho(r) &\geq \int_{-\tilde{K}}^{\tilde{K}} \rho(t) \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(r-t)^2}{2}\right) dt \\
&\geq \frac{1}{\sqrt{2\pi}} \mathbf{P}(|X| \leq \tilde{K}) \cdot \min_{0 \leq r \leq \tilde{K}, t \in [-\tilde{K}, \tilde{K}]} \exp\left(-\frac{(r-t)^2}{2}\right) = \frac{1}{2\sqrt{2\pi}} \exp\left(-2\tilde{K}^2\right).
\end{aligned}
$$

We further notice that $1 - F(r) \leq 1$. Hence for any $\varepsilon > 0$, if choosing $M_1 = \left(\frac{1}{2\sqrt{2\pi}} \exp\left(-2\tilde{K}^2\right)\right)^{-\beta+\epsilon}$, we will have

$$
1 - F(r) \leq 1 \leq M_1 \rho(r)^{\beta-\epsilon}, \quad \forall r \in [0, R_0].
$$

Next, we consider cases where $r > \tilde{K}$. According to the assumption, we have

$$
\mathbf{P}(X \geq r) \leq C \exp\left(-\frac{r^2}{2K^2}\right),
$$

which indicates that

$$\begin{aligned}1-F(r) &= \int_{-\infty}^{r} \eta(t)(1-\Phi_1(r-t))dt + \int_{r}^{\infty} \eta(t)(1-\Phi_1(r-t))dt \\ &\leq \int_{-\infty}^{r} \eta(t)(1-\Phi_1(r-t))dt + \int_{r}^{\infty} \eta(t)dt \\ &\leq \int_{-\infty}^{r} \eta(t)(1-\Phi_1(r-t))dt + \mathbf{P}(X \geq r).\end{aligned}$$

For $r > t$, according to Lemma 3, we have $1-\Phi_1(r-t) \leq \exp\left(-\frac{(r-t)^2}{2}\right)$, which indicates that

$$\begin{aligned}1-F(r) &\leq \int_{-\infty}^{r} \eta(t)\exp\left(-\frac{(r-t)^2}{2}\right)dt + \mathbf{P}(X \geq r) \\ &\leq \int_{-\infty}^{\infty} \eta(t)\exp\left(-\frac{(r-t)^2}{2}\right)dt + \mathbf{P}(X \geq r) = \sqrt{2\pi}\cdot\rho(r) + \mathbf{P}(X \geq r).\end{aligned}$$

Moreover, according to Lemma 4, we know that there exists constant C_ϵ such that

$$\rho(r) \geq C_\epsilon \mathbf{P}[X \geq r]^{\frac{1}{\beta-\epsilon}},$$

which indicates that $\mathbf{P}(X \geq r) \leq C_\epsilon^{-\beta+\epsilon}\rho(r)^{\beta-\epsilon}$, which indicates that

$$1-F(r) \leq \sqrt{2\pi}\cdot\rho(r) + C_\epsilon^{-\beta+\epsilon}\rho(r)^{\beta-\epsilon}.$$

When $\rho(r) \leq 1$, since $\beta \in [0,1]$, we have

$$\sqrt{2\pi}\cdot\rho(r) \leq \sqrt{2\pi}\cdot\rho(r)^{\beta-\epsilon}.$$

Therefore,

$$1-F(r) \leq (C_\epsilon^{-\beta+\epsilon} + \sqrt{2\pi})\cdot\rho(r)^{\beta-\epsilon}.$$

When $\rho(r) > 1$, we will also have $\rho(r)^{\beta-\epsilon} > 1$. Hence the following inequality holds

$$1-F(r) \leq 1 < \rho(r)^{\beta-\epsilon}.$$

Above all, if we choose $M = \max\{M_1, (C_\epsilon^{-\beta+\epsilon} + \sqrt{2\pi}), 1\} \geq 1$, then we have

$$1-F(r) \leq M\rho(r)^{\beta-\epsilon}, \quad \forall r \geq 0.$$

$\square$

Proposition 8. *Consider two distributions $\mathbb{P}, \mathbb{Q}$ on $\mathbb{R}$. Assume that the distribution $\mathbb{Q}$ has a strictly positive PDF. We denote the PDF of $\mathbb{P}$ as $\rho_{\mathbb{P}}(\cdot)$, and the CDFs of $\mathbb{P}, \mathbb{Q}$ as $F_{\mathbb{P}}, F_{\mathbb{Q}}$, respectively. For a fixed $h > 0$ denote $L_h(t) \triangleq \sup_{x\in[t-h,t+h]} |F_{\mathbb{P}}(x) - F_{\mathbb{Q}}(x)|$ and $\underline{\rho}_h(t) = \inf_{x\in[t-h,t+h]} \rho_{\mathbb{Q}}(x)$. If we have*

$$\Delta_h(t) \triangleq \frac{L_h(t)}{\underline{\rho}_h(t)} \leq h,$$

then

$$\left|F_{\mathbb{Q}}^{-1}(F_{\mathbb{P}}(t)) - t\right| \leq \Delta_h(t)\,.$$

Proof. Suppose that $F_{\mathbb{P}}(t) > F_{\mathbb{Q}}(t)$ and let $h' = \Delta_h(t) \leq h$. Then we claim that

$$F_{\mathbb{P}}(t) \leq F_{\mathbb{Q}}(t+h')\,. \tag{52}$$

Indeed, we have $F_{\mathbb{P}}(t+h') \geq F_{\mathbb{P}}(t) + \underline{\rho}_h(t)h' = F_{\mathbb{P}}(t) + L_h(t)$. On the other hand, $F_{\mathbb{P}}(t+h') \leq F_{\mathbb{Q}}(t+h') + L_h(t)$. Combining these two, we obtain (52). Now, since $F_{\mathbb{Q}}(t) < F_{\mathbb{P}}(t) \leq F_{\mathbb{Q}}(t+h')$ we obtain that $0 < F_{\mathbb{Q}}^{-1}(F_{\mathbb{P}}(t)) - t \leq h'$. The case of $F_{\mathbb{P}}(t) < F_{\mathbb{Q}}(t)$ is treated similarly. $\square$

Proposition 9. *Assume distribution $\mathbb{P}, \mathbb{Q}$ are K_1, K_2-subgaussian distributions respectively, e.g. for any $X \sim \mathbb{P}, Y \sim \mathbb{Q}$ we have*

$$\mathbf{P}(|X| \geq r) \leq C_1 \exp\left(-\frac{r^2}{2K_1^2}\right), \quad \mathbf{P}(|Y| \geq r) \leq C_2 \exp\left(-\frac{r^2}{2K_2^2}\right), \quad \forall r > 0\,.$$

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

$$\left|F_{\mathbb{Q},1}^{-1}(F_{\mathbb{P},1}(x)) - x\right| \leq 2|x| + 2 + \tilde{K}_1 + \tilde{K}_2(|x| + 2 + \tilde{K}_1)\,,$$

where $F_{\mathbb{P},1}, F_{\mathbb{Q},1}$ denote the CDFs of $X + Z$ and $Y + Z$, $(X, Y) \perp\!\!\!\perp Z \sim \mathcal{N}$, and $\tilde{K}_1 \triangleq K_1\sqrt{2\log 2C_1}, \tilde{K}_2(t) \triangleq K_2 t + K_2\sqrt{2\log(4tC_2)}$.

Proof. We let $R = |x|$ and $\tilde{R} = R + 2 + \tilde{K}_1$. First we notice that the PDFs of distribution $\mathbb{P} * \mathcal{N}, \mathbb{Q} * \mathcal{N}$ at any real number is positive, hence $F_{\mathbb{P},1}, F_{\mathbb{Q},1}$ are monotonically increasing in the entire real line. We have

$$\mathbf{P}\left(|X| \geq \tilde{K}_1\right) \leq C_1 \exp\left(-\log(2C_1)\right) = \frac{1}{2}.$$

Therefore, we obtain that

$$\mathbf{P}\left(|X| \leq \tilde{K}_1\right) \geq 1 - \frac{1}{2} = \frac{1}{2}.$$

We further notice that if $X \sim \mathbb{P}, Z \sim \mathcal{N}(0,1)$ are independent, $X + Z \sim \mathbb{P} * \mathcal{N}(0,1)$. And also

$$\begin{aligned}\{|X| \leq \tilde{K}_1\} \cap \{Z \leq -\tilde{K}_1 - R\} &\subset \{X + Z \leq -R\}\\ \{|X| \leq \tilde{K}_1\} \cap \{Z \geq \tilde{K}_1 + R\} &\subset \{X + Z \geq R\}\,.\end{aligned}$$

Recall that we use Φ_1 to denote the CDF of distribution $\mathcal{N}(0,1)$. Hence noticing that $\Phi_1(-R - \tilde{K}_1) = 1 - \Phi_1(R + \tilde{K}_1) = \mathbf{P}(Z \leq -\tilde{K}_1 - R) = \mathbb{P}(Z \geq \tilde{K}_1 + R)$, we have

$$\begin{aligned}\frac{1}{2}\Phi_1(-R - \tilde{K}_1) &\leq \mathbf{P}\left(|X| \geq \tilde{K}_1\right)\mathbf{P}(Z \leq -\tilde{K}_1 - R) \leq \mathbf{P}(X + Z \leq -R) = F_{\mathbb{P},1}(-R)\\ \frac{1}{2}\Phi_1(-R - \tilde{K}_1) &\leq \mathbf{P}\left(|X| \geq \tilde{K}_1\right)\mathbf{P}(Z \geq \tilde{K}_1 + R) \leq \mathbf{P}(X + Z \geq R) = 1 - F_{\mathbb{P},1}(R),\end{aligned}$$

which indicates that

$$\frac{1}{2}\Phi_1(-R - \tilde{K}_1) \leq F_{\mathbb{P},1}(-R) \leq F_{\mathbb{P},1}(R) \leq 1 - \frac{1}{2}\Phi_1(-R - \tilde{K}_1).$$

Next, if $Y \sim \mathbb{Q}, Z \sim \mathcal{N}(0,1)$ are independent, we have $Y + Z \sim \mathbb{Q} * \mathcal{N}(0,1)$. Noticing that,

$$\begin{aligned}\left\{Y + Z \leq -\tilde{R} - \tilde{K}_2(\tilde{R})\right\} &\subset \{Z \leq -\tilde{R}\} \cup \{Y \leq -\tilde{K}_2(\tilde{R})\},\\ \left\{Y + Z \geq \tilde{R} + \tilde{K}_2(\tilde{R})\right\} &\subset \{Z \geq \tilde{R}\} \cup \{Y \geq \tilde{K}_2(\tilde{R})\},\end{aligned}$$

we obtain that

$$\begin{aligned}F_{\mathbb{Q},1}(-\tilde{R} - \tilde{K}_2(\tilde{R})) &\leq \Phi_1(-\tilde{R}) + \mathbf{P}(|Y| \geq \tilde{K}_2(\tilde{R})),\\ 1 - F_{\mathbb{Q},1}(\tilde{R} + \tilde{K}_2(\tilde{R})) &\leq 1 - \Phi_1(\tilde{R}) + \mathbf{P}(|Y| \geq \tilde{K}_2(\tilde{R})) = \Phi_1(-\tilde{R}) + \mathbf{P}(|Y| \geq \tilde{K}_2(\tilde{R})).\end{aligned}$$

According to Proposition 2.1.2 in [Ver18], we have

$$\Phi_1(-\tilde{R}) \geq \left(\frac{1}{\tilde{R}} - \frac{1}{\tilde{R}^3}\right) \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\tilde{R}^2}{2}\right).$$

Hence since $\tilde{R} = R + \tilde{K}_1 + 2 \geq 2$, we will have

$$\Phi_1(-\tilde{R}) \geq \frac{3}{4\tilde{R}} \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\tilde{R}^2}{2}\right) \geq \frac{1}{4\tilde{R}} \exp\left(-\frac{\tilde{R}^2}{2}\right).$$

We further notice that $\tilde{K}_2(\tilde{R}) = K_2\tilde{R} + K_2\sqrt{2\log\left(4\tilde{R}C_2\right)}$, hence we have

$$\mathbf{P}(|Y| \geq \tilde{K}_2(\tilde{R})) \leq C_2 \exp\left(-\frac{\tilde{K}_2(\tilde{R})^2}{2K_2^2}\right) = \frac{1}{4\tilde{R}} \exp\left(-\frac{\tilde{R}^2}{2}\right) \leq \Phi_1(-\tilde{R}),$$

which indicates that

$$F_{\mathbb{Q},1}(-\tilde{R} - \tilde{K}_2(\tilde{R})) \leq 2\Phi_1(-\tilde{R}), \quad 1 - F_{\mathbb{Q},1}(\tilde{R} + \tilde{K}_2(\tilde{R})) \leq 2\Phi_1(-\tilde{R}).$$

Additionally, since for any $t \leq 0$, we have

$$\exp\left(-\frac{(t-2)^2}{2}\right) \leq \exp\left(-\frac{t^2}{2} - \frac{4}{2}\right) = \exp(-2) \cdot \exp\left(-\frac{t^2}{2}\right) \leq \frac{1}{4}\exp\left(-\frac{t^2}{2}\right).$$

This indicates that

$$\begin{aligned}
\frac{1}{4}\Phi_1(-R - \tilde{K}_1) &= \frac{1}{4} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-R-\tilde{K}_1} \exp\left(-\frac{t^2}{2}\right) dt \\
&\geq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-R-\tilde{K}_1} \exp\left(-\frac{(t-2)^2}{2}\right) dt \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-R-\tilde{K}_1-2} \exp\left(-\frac{t^2}{2}\right) dt = \Phi_1(-R - \tilde{K}_1 - 2).
\end{aligned}$$

Therefore, we obtain that

$$F_{\mathbb{Q},1}\left(-\tilde{R} - \tilde{K}_2(\tilde{R})\right) \leq 2\Phi_1\left(-\tilde{R}\right) = 2\Phi_1(-R - \tilde{K}_1 - 2) \leq \frac{1}{2}\Phi_1(-R - \tilde{K}_1) \leq F_{\mathbb{P},1}(-R).$$

Similarly, we can also obtain that

$$F_{\mathbb{P},1}(R) \leq F_{\mathbb{Q},1}\left(\tilde{R} + \tilde{K}_2(\tilde{R})\right).$$

Hence using the monotonicity of $F_{\mathbb{P},1}$ and $F_{\mathbb{Q},1}$, we obtain that,

$$F_{\mathbb{Q},1}\left(-\tilde{R} - \tilde{K}_2(\tilde{R})\right) \leq F_{\mathbb{P},1}(-R) \leq F_{\mathbb{P},1}(x) \leq F_{\mathbb{P},1}(R) \leq F_{\mathbb{Q},1}\left(\tilde{R} + \tilde{K}_2(\tilde{R})\right),$$

which indicates that

$$-\tilde{R} - \tilde{K}_2(\tilde{R}) \leq F_{\mathbb{Q},1}^{-1}(F_{\mathbb{P},1}(x)) \leq \tilde{R} + \tilde{K}_2(\tilde{R}).$$

Hence we have

$$\left|F_{\mathbb{Q},1}^{-1}(F_{\mathbb{P},1}(x)) - x\right| \leq R + \tilde{R} + \tilde{K}_2(\tilde{R}) = 2|x| + 2 + \tilde{K}_1 + \tilde{K}_2(|x| + \tilde{K}_1 + 2).$$

□

Proposition 10. *Suppose F, F_n are CDFs of distribution $\mathbb{P} * \mathcal{N}$ and $\mathbb{P}_n * \mathcal{N}$ respectively. Then with probability at least $1 - \delta$, we have the following inequality:*

$$\sup_{t \in \mathbb{R}} \frac{|F(t) - F_n(t)|}{\sqrt{1/n \vee (F(t) \wedge (1 - F(t)))}} \leq \frac{16}{\sqrt{n}} \log\left(\frac{2n}{\delta}\right).$$

To prove this proposition, we first present a lemma indicating a similar result without Gaussian smoothing:

Lemma 5. *For a given distribution $\mathbb{Q}$ on real numbers with always-positive PDF, we denote its empirical measure with n data points to be $\mathbb{Q}_n$ ($\mathbb{Q}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ where $X_i \sim \mathbb{Q}$ are i.i.d.). We further use $F, \hat{F}_n$ to denote the CDF of $\mathbb{Q}, \mathbb{Q}_n$ respectively.Then with probability at least $1-\delta$, we have*

$$\sup_{t\in\mathbb{R}} \frac{|F(t)-\hat{F}_n(t)|}{\sqrt{1/n \vee (F(t)\wedge(1-F(t)))}} \leq \frac{8}{\sqrt{n}}\log\left(\frac{n}{\delta}\right).$$

Remark 5. *From Theorem 2.1 of [GK+06] we can obtain a result similar to this lemma: if $\mathbb{Q}$ is the uniform distribution on $[0,1]$, then there exist universal positive constants C_0, K such that for any $s>0$,*

$$\mathbf{P}\left[\sup_{1/n\leq t\leq 1/2} \frac{|F(t)-\hat{F}_n(t)|}{\sqrt{t}} \geq \frac{4}{\sqrt{n}} + \frac{2s\log\log n}{\sqrt{n}\log\log\log n}\right] \leq \frac{C_0}{\log(n)^{(s/(2K)-1)}}.$$

Remark 6. *If we would like to obtain a uniform bound without truncation, then we have to pay an additional factor $\sqrt{1/\delta}$. This is summarized in the following results: with probability at least $1-\delta$, we have*

$$\sup_{t\in\mathbb{R}} \frac{|F(t)-\hat{F}_n(t)|}{\sqrt{F(t)\wedge(1-F(t))}} \leq 16\sqrt{\frac{1}{\delta n}}\log\left(\frac{4n}{\delta}\right).$$

Also we have a lower bound to the LHS in the above inequality, indicating that the factor $\sqrt{1/\delta}$ is necessary: with probability at least δ, we have

$$\sup_{t\in\mathbb{R}} \frac{|F(t)-\hat{F}_n(t)|}{\sqrt{F(t)\wedge(1-F(t))}} \geq \sqrt{\frac{1}{2\delta n}}.$$

Proof of Lemma 5. With loss of generality, we assume $\mathbb{Q}$ is the uniform distribution on $[0,1]$ (otherwise we consider the similar argument on distribution $\mathbb{Q}(F^{-1}(\cdot))$). Then we have $F(t)=t$ for any $0\leq t\leq 1$. We only need to prove that with probability at least $1-\delta$,

$$\sup_{t\in\mathbb{R}} \frac{|F(t)-\hat{F}_n(t)|}{\sqrt{1/n \vee (t\wedge(1-t))}} \leq \sqrt{\frac{\log n}{n}}.$$

According to Bernstein inequality, we have

$$\mathbf{P}\left(\left|F(t)-\hat{F}_n(t)\right| > \varepsilon\right) \leq \exp\left(-\frac{n\varepsilon^2}{2t(1-t)+2/3\varepsilon}\right) \leq \exp\left(-\frac{n\varepsilon^2}{2t+2/3\varepsilon}\right).$$

Choosing $\varepsilon = 4\sqrt{\frac{t}{n}}\log\left(\frac{1}{\delta}\right)$, and noticing that with this choice we have $\frac{1}{2}n\varepsilon^2 \geq 2t\log(1/\delta)$ and also $\frac{1}{2}n\varepsilon^2 \geq \frac{2}{3}\varepsilon\log(1/\delta)$, we obtain that

$$\mathbf{P}\left(\left|F(t)-\hat{F}_n(t)\right| > 4\sqrt{\frac{t}{n}}\log\left(\frac{1}{\delta}\right)\right) \leq \delta.$$

Therefore, choosing $t=\frac{k}{n}$ with $1\leq k\leq \frac{n}{2}$, and applying union bound for $1\leq k\leq\frac{n}{2}$, we obtain that

$$\mathbf{P}\left(\left|F\left(\frac{k}{n}\right)-\hat{F}_n\left(\frac{k}{n}\right)\right| \leq 4\sqrt{\frac{(k/n)}{n}}\log\left(\frac{n}{\delta}\right),\ \forall 1\leq k\leq\frac{n}{2}\right) \leq \frac{\delta}{2}.$$

We further notice that for any $\frac{k}{n}\leq t\leq\frac{k+1}{n}$, we have

$$|F(t)-\hat{F}_n(t)| = |t-\hat{F}_n(t)| \leq \frac{1}{n} + \max\left\{\left|F\left(\frac{k}{n}\right)-\hat{F}_n\left(\frac{k}{n}\right)\right|, \left|F\left(\frac{k+1}{n}\right)-\hat{F}_n\left(\frac{k+1}{n}\right)\right|\right\}.$$

When $k \geq 1$ and $\frac{2k}{n} \leq \frac{k+1}{n}$. Therefore, if for every $1 \leq k \leq \frac{n}{2}$ we all have $\left|F\left(\frac{k}{n}\right) - \hat{F}_n\left(\frac{k}{n}\right)\right| \leq 4\sqrt{\frac{(k/n)}{n}} \log\left(\frac{n}{\delta}\right)$, then for every $0 \leq t \leq \frac{1}{n}$, we have

$$\frac{|F(t) - \hat{F}_n(t)|}{\sqrt{1/n \vee (t \wedge (1-t))}} \leq \frac{1/n + |F(1/n) - \hat{F}_n(1/n)|}{\sqrt{1/n}} \leq 5\sqrt{\frac{1}{n}} \log\left(\frac{n}{\delta}\right),$$

and for every $\frac{k}{n} \leq t \leq \frac{k+1}{n}$ with $k \leq \frac{n}{2}$, we have

$$\begin{aligned}
\frac{|F(t) - \hat{F}_n(t)|}{\sqrt{1/n \vee (t \wedge (1-t))}} &\leq \frac{\frac{1}{n} + \max\left\{\left|F\left(\frac{k}{n}\right) - \hat{F}_n\left(\frac{k}{n}\right)\right|, \left|F\left(\frac{k+1}{n}\right) - \hat{F}_n\left(\frac{k+1}{n}\right)\right|\right\}}{\sqrt{k/n}} \\
&\leq \sqrt{\frac{1}{n}} + \sqrt{2} \cdot \max\left\{\frac{\left|F\left(\frac{k}{n}\right) - \hat{F}_n\left(\frac{k}{n}\right)\right|}{\sqrt{k/n}}, \frac{\left|F\left(\frac{k+1}{n}\right) - \hat{F}_n\left(\frac{k+1}{n}\right)\right|}{\sqrt{(k+1)/n}}\right\} \\
&\leq \sqrt{\frac{1}{n}} + 4\sqrt{2} \cdot \sqrt{\frac{1}{n}} \log\left(\frac{n}{\delta}\right) \leq 8\sqrt{\frac{1}{n}} \log\left(\frac{n}{\delta}\right).
\end{aligned}$$

Therefore, we have proved that with probability at least $1 - \frac{\delta}{2}$,

$$\frac{|F(t) - \hat{F}_n(t)|}{\sqrt{1/n \vee (t \wedge (1-t))}} \leq 8\sqrt{\frac{1}{n}} \log\left(\frac{n}{\delta}\right)$$

holds for every $0 \leq t \leq \frac{1}{2}$. Similarly, we can prove that with probability at least $1 - \frac{\delta}{2}$, the above inequality holds for $\frac{1}{2} \leq t \leq 1$. Therefore, with probability at least $1 - \delta$, we have

$$\sup_{0 \leq t \leq 1} \frac{|F(t) - \hat{F}_n(t)|}{\sqrt{1/n \vee (t \wedge (1-t))}} \leq 8\sqrt{\frac{1}{n}} \log\left(\frac{n}{\delta}\right).$$

This completes the proof of this lemma.

□

Proof of Proposition 10. Suppose random variables $X \sim \mathbb{P}, Y \sim \mathcal{N}$ are independent. Then $X + Y \sim \mathbb{P} * \mathcal{N}$. We generate n *i.i.d.* samples $X_1, \cdots, X_n; Y_1, \cdots, Y_n$. Then $X_i + Y_i$ are n *i.i.d.* samples of $\mathbb{P} * \mathcal{N}$. We use $\hat{F}_n$ to denote the PDF of empirical measure $\hat{\mathbb{P}}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i + Y_i}$. Then according to Lemma 5, we have with probability $1 - \delta$,

$$\sup_{t \in \mathbb{R}} \frac{|F(t) - \hat{F}_n(t)|}{\sqrt{1/n \vee (F(t) \wedge (1 - F(t)))}} \leq \frac{8}{\sqrt{n}} \log\left(\frac{n}{\delta}\right).$$

Hence Markov inequality indicates that

$$\mathbf{P}\left(\exp\left(\sup_{t \in \mathbb{R}} \frac{\sqrt{n}}{16} \cdot \frac{|F(t) - \hat{F}_n(t)|}{\sqrt{1/n \vee (F(t) \wedge (1 - F(t)))}} - \frac{\log(n)}{2}\right) \geq \frac{1}{\delta}\right) \leq \delta^2.$$

Therefore, we have

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\sup_{t \in \mathbb{R}} \frac{\sqrt{n}}{16} \cdot \frac{|F(t) - \hat{F}_n(t)|}{\sqrt{1/n \vee (F(t) \wedge (1 - F(t)))}} - \frac{\log(n)}{2}\right)\right] \\
&= 1 + \int_1^\infty \mathbf{P}\left(\exp\left(\sup_{t \in \mathbb{R}} \frac{\sqrt{n}}{16} \cdot \frac{|F(t) - \hat{F}_n(t)|}{\sqrt{1/n \vee (F(t) \wedge (1 - F(t)))}} - \frac{\log(n)}{2}\right) \geq r\right) dr \\
&\leq 1 + \int_1^\infty \frac{1}{r^2} dr \\
&= 2.
\end{aligned}$$

Moreover, we notice that

$$\mathbb{E}\left[\hat{F}_n(t)\Big|X_1,\cdots,X_n\right] = \mathbf{P}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i+Y_i)\le t\Big|X_1,\cdots,X_n\right) = F_n(t),$$

where F_n is the CDF of $\mathbb{P}_n * \mathcal{N}$ with $\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$. Hence according to the Jensen's inequality and the convexity of function $|\cdot|$ and $\exp(\cdot)$, we have

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\sup_{t\in\mathbb{R}}\frac{\sqrt{n}}{16}\cdot\frac{|F(t)-F_n(t)|}{\sqrt{1/n\vee(F(t)\wedge(1-F(t)))}}-\log(n)\right)\right]\\
&\le \mathbb{E}\left[\exp\left(\sup_{t\in\mathbb{R}}\frac{\sqrt{n}}{16}\cdot\frac{|F(t)-F_n(t)|}{\sqrt{1/n\vee(F(t)\wedge(1-F(t)))}}-\frac{\log(n)}{2}\right)\right]\\
&= \mathbb{E}\left[\exp\left(\sup_{t\in\mathbb{R}}\frac{\sqrt{n}}{16}\cdot\frac{|F(t)-\mathbb{E}_{Y_i,1\le i\le n}[\hat{F}_n(t)]|}{\sqrt{1/n\vee(F(t)\wedge(1-F(t)))}}-\frac{\log(n)}{2}\right)\right]\\
&\le \mathbb{E}\left[\exp\left(\sup_{t\in\mathbb{R}}\frac{\sqrt{n}}{16}\cdot\frac{\mathbb{E}_{Y_i,1\le i\le n}|F(t)-\hat{F}_n(t)|}{\sqrt{1/n\vee(F(t)\wedge(1-F(t)))}}-\frac{\log(n)}{2}\right)\right]\\
&\le \mathbb{E}\left[\exp\left(\mathbb{E}\left[\sup_{t\in\mathbb{R}}\frac{\sqrt{n}}{16}\cdot\frac{|F(t)-\hat{F}_n(t)|}{\sqrt{1/n\vee(F(t)\wedge(1-F(t)))}}-\frac{\log(n)}{2}\Bigg|X_1,\cdots,X_n\right]\right)\right]\\
&\le \mathbb{E}\left[\mathbb{E}\left[\exp\left(\sup_{t\in\mathbb{R}}\frac{\sqrt{n}}{16}\cdot\frac{|F(t)-\hat{F}_n(t)|}{\sqrt{1/n\vee(F(t)\wedge(1-F(t)))}}-\frac{\log(n)}{2}\right)\right]\Bigg|X_1,\cdots,X_n\right]\\
&\le 2.
\end{aligned}$$

And according to Markov inequality, we have

$$\mathbf{P}\left(\exp\left(\sup_{t\in\mathbb{R}}\frac{\sqrt{n}}{16}\cdot\frac{|F(t)-F_n(t)|}{\sqrt{1/n\vee(F(t)\wedge(1-F(t)))}}-\log(n)\right)\ge\frac{2}{\delta}\right)\le\delta.$$

Therefore, with probability at least $1-\delta$ we have

$$\sup_{t\in\mathbb{R}}\frac{|F_n(t)-F(t)|}{\sqrt{1/n\vee(F(t)\wedge(1-F(t)))}}\le\frac{16}{\sqrt{n}}\log\left(\frac{2n}{\delta}\right).$$

□

We are now ready to prove the upper bound part of Theorem 2

Proof of the Upper Bound in Theorem 2. In the following proof, we use $\mathcal{N}$ to denote the 1-dimensional standard normal distribution $\mathcal{N}$, and $T(\cdot)$ to denote the push-forward operator between $\mathbb{P}*\mathcal{N}$ and $\mathbb{P}_n*\mathcal{N}$ ($T(t) = F_n^{-1}(F(t))$, where F, F_n are CDF of distribution $\mathbb{P}*\mathcal{N}$ and $\mathbb{P}_n*\mathcal{N}$ respectively).

First, as shown in (41) in the outline of the proof, we show that if E is any event of probability at least $1-\frac{C_E}{n^2}$ for some constant C_E only depending on C, K, then we have

$$\mathbb{E}\left[W_2^2(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})\right]\le\mathbb{E}\left[W_2^2(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})|E\right]+\mathcal{O}\left(\frac{1}{n}\right). \tag{53}$$

Actually we have

$$\mathbb{E}\left[W_2^2(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})\right]=\mathbb{E}\left[W_2^2(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})\mathbf{1}_E\right]+\mathbb{E}\left[W_2^2(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})\mathbf{1}_{E^c}\right],$$

and $\mathbb{E}\left[W_2^2(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})\mathbf{1}_E\right] = \mathbb{E}\left[W_2^2(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})|E\right]\mathbf{P}[E] \le \mathbb{E}\left[W_2^2(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})|E\right]$. As for the second term, according to Cauchy-Schwartz inequality we have

$$\begin{aligned}
\mathbb{E}\left[W_2^2(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})\mathbf{1}_{E^c}\right] &\le \sqrt{\mathbb{E}[\mathbf{1}_E]\mathbb{E}[W_2^4(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})]} = \sqrt{\mathbf{P}[E^c]\mathbb{E}[W_2^4(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})]} \\
&\le \frac{\sqrt{C_E}}{n}\sqrt{\mathbb{E}[W_2^4(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})]}.
\end{aligned}$$

We further notice that according to the triangle inequality of W2 distance we have

$$\begin{aligned}
\mathbb{E}[W_2^4(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})] &\le \mathbb{E}[(W_2(\mathbb{P}*\mathcal{N},\delta_0)+W_2(\delta_0,\mathbb{P}_n*\mathcal{N}))^4] \le \mathbb{E}[8W_2(\mathbb{P}*\mathcal{N},\delta_0)^4+8W_2(\delta_0,\mathbb{P}_n*\mathcal{N})^4] \\
&= 8\mathbb{E}[(V_1+Z)^4]+8\mathbb{E}[\mathbb{E}[(V_2+Z)^4|X_{1:n}]] = 64\mathbb{E}[V_1^4]+64\mathbb{E}[\mathbb{E}[V_2^4|X_{1:n}]]+128\mathbb{E}[Z^4] \\
&= \mathcal{O}(1),
\end{aligned}$$

where we use δ_0 to denote the delta distribution at 0, and $V_1\sim\mathbb{P}, V_2\sim\mathbb{P}_n, Z\sim\mathcal{N}$ are all independent. The last equation is because $\mathbb{P}$ is K-subgaussian, hence all moments of $\mathbb{P}$ are upper bounded by constant. Hence we have proved (53).

Next, we notice that for any n *i.i.d.* samples $X_1,\cdots,X_n$, we have

$$\mathbf{P}\left[\left\{|X_i|\le 2K\sqrt{2\log n},\forall 1\le i\le n\right\}\right] = \left(1-\mathbb{P}\left[|X_1|\ge 2K\sqrt{2\log n}\right]\right)^n \ge \left(1-\frac{C}{n^4}\right)^n \ge 1-\frac{C}{n^3}\ge 1-\frac{C}{n^2},$$

where we use the fact that $\mathbb{P}$ is a K-subgaussian distribution ($\mathbf{P}[|X|\ge t]\le C\exp(-t^2/(2K^2))$). Therefore, with probability at least $1-\frac{C}{n^2}$ we have for $X'\sim\mathbb{P}_n$

$$\mathbf{P}(|X'|\ge r)) \le e\exp\left(-\frac{r^2}{2(2K\sqrt{\log n})^2}\right),$$

which indicates that $\mathbb{P}_n$ is $2K\sqrt{\log n}$ subgaussian with probability at least $1-\frac{C}{n^2}$. We assume this event to be E, then $\mathbf{P}[E]\ge 1-\frac{C}{n^2}$. In the following proof we all assume the event E, where we will deal with E^c in the end. Noticing that for $X\sim\mathbb{P}, Y\sim\mathbb{P}_n$ we have

$$\mathbf{P}[|X|\ge r]\le C\exp\left(-\frac{r^2}{2K^2}\right), \quad \mathbf{P}[|Y|\ge r]\le e\exp\left(-\frac{r^2}{2(2K\sqrt{\log n})^2}\right),$$

according to Proposition 9 we obtain that

$$\begin{aligned}
|T(x)-x| &\le 2|x|+2+K\sqrt{2\log(2C)}+(2K\sqrt{\log n})(|x|+2+K+\sqrt{2\log(4e(|x|+2+K))}) && (54)\\
&\le C_1+C_2\sqrt{\log n}|x| && (55)
\end{aligned}$$

for some positive constant C_1, C_2 only depending on C, K. We further notice that according to (48) we have

$$\rho(x)\le\left(\frac{1}{\sqrt{2\pi}}+C\right)\exp\left(-\frac{x^2}{8K^2}\right).$$

Hence when $|x|\ge 2K\sqrt{2\log n}$, we will have $\exp\left(-\frac{x^2}{8K^2}\right)\le\frac{1}{n}$ and hence

$$\int_{|t|>2K\sqrt{2\log n}}\rho(t)|T(t)-t|^2 \le \left(\frac{1}{\sqrt{2\pi}}+C\right)\int_{|t|>2K\sqrt{2\log n}}\exp\left(-\frac{t^2}{8K^2}\right)(C_1+C_2\sqrt{\log n}|t|)^2dt = \tilde{\mathcal{O}}\left(\frac{1}{n}\right)$$

Therefore, we only need to analyze the integral

$$\int_{|t|\le 2K\sqrt{2\log n}}\rho(t)|T(t)-t|^2. \tag{56}$$

In what follows, we will use the notation:

$$\overline{\rho}(t) = \sup_{x\in[t-1,t+1]} \rho(x), \quad \underline{\rho}(t) = \inf_{x\in[t-1,t+1]} \rho(x)$$
$$\Lambda(t) = \sup_{x\in[t-1,t+1]} |F(x) - F_n(x)|.$$

The key idea to bound the integral in (56) is the following observation from Proposition 8: if $\Lambda(t) \le \underline{\rho}(t)$, then we have

$$|T(t) - t| \le \frac{\Lambda(t)}{\underline{\rho}(t)},$$

which indicates that

$$\rho(t)|T(t) - t|^2 \le \frac{\Lambda(t)^2}{\rho(t)} \cdot \left(\frac{\rho(T)}{\underline{\rho}(t)}\right)^2.$$

In the following proof, we will use the concentration proposition (Proposition 10) to divide the interval $[-2K\sqrt{\log n}, 2K\sqrt{\log n}]$ into the set where $\Lambda(t) \le \underline{\rho}(t)$ where the integral can be bounded from the above inequality, and the set where $\rho(t)$ is very small hence $\rho(t)|T(t)-t|^2$ won't have much effect in the integral.

According to Proposition 10, with probability at least $1 - \frac{1}{n^2}$ we have

$$\sup_{t\in\mathbb{R}} \frac{|F(t) - F_n(t)|}{\sqrt{1/n \vee \min\{F(t), 1-F(t)\}}} \le \frac{16}{\sqrt{n}} \log\left(2n^3\right).$$

We assume this event to be E_1, where $\mathbf{P}[E_1] \ge 1 - \frac{1}{n^2}$. In the rest of the proof we assume E_1 and will deal with E_1^c in the end. Then we have

$$\begin{aligned}\Lambda(t) = \sup_{x\in[t-1,t+1]} |F(x) - F_n(x)| &\le \frac{16\log(2n^3)}{\sqrt{n}} \sup_{x\in[t-1,t+1]} \sqrt{\frac{1}{n} \vee \min\{F(x), 1-F(x)\}} \\ &= \frac{16\log(2n^3)}{\sqrt{n}} \sqrt{\frac{1}{n} \vee \sup_{x\in[t-1,t+1]} \min\{F(x), 1-F(x)\}}.\end{aligned}$$

According to Proposition 7, for any $0 < \epsilon < \beta$, $\exists M = M(K, C, \epsilon) \ge 1$ such that

$$\min\{F(x), 1-F(x)\} \le M\rho(r)^{\beta-\epsilon},$$

which indicates that

$$\Lambda(t) \le \frac{16\log(2n^3)}{\sqrt{n}} \sqrt{\frac{1}{n} \vee \sup_{x\in[t-1,t+1]} M\rho(x)^{\beta-\epsilon}} = \frac{16\log(2n^3)}{n} \vee \frac{16\sqrt{M}\log(2n^3)}{\sqrt{n}} \overline{\rho}(t)^{\frac{\beta-\epsilon}{2}}$$

Next we will upper bound $\frac{\overline{\rho}(t)}{\rho(t)}$ and also $\frac{\rho(t)}{\underline{\rho}(t)}$ from the following observation: Noticing that for $S \sim \mathbb{P}$ we have

$$\mathbb{E}[S] = \int_{-\infty}^{\infty} x\eta(x)dx \le \int_{-\infty}^{\infty} |x|\eta(x)dx = \int_0^{\infty} \mathbf{P}[S \ge r]dr \le \int_0^{\infty} C\exp\left(-\frac{r^2}{2K^2}\right)dr = \frac{CK\sqrt{2\pi}}{2} \le 2CK,$$

hence according to [PW16, Prop. 2], we obtain that $\mathbb{P} * \mathcal{N}$ is $(3, 8CK)$-regular, which indicates that for $|t| \le 2K\sqrt{2\log n}$ and $\forall x \in [t-1, t+1]$,

$$\begin{aligned}\frac{\rho(x)}{\rho(t)} &\le \exp\left(3(|t|+1) + 8CK\right) \le \exp\left(6K\sqrt{2\log n} + 3 + 8CK\right) \triangleq L(n) \\ \frac{\rho(x)}{\rho(t)} &\ge \exp\left(-3(|t|+1) - 8CK\right) \le \exp\left(-6K\sqrt{2\log n} - 3 - 8CK\right) = \frac{1}{L(n)}.\end{aligned}$$

Hence we have $1 \le \frac{\overline{\rho}(t)}{\rho(t)}, \frac{\rho(t)}{\underline{\rho}(t)} \le L(n)$. Therefore, when

$$\rho(t) \ge \frac{16\log(2n^3)L(n)}{n} \vee \left(\frac{256M\log^2(2n^3)}{n}L(n)^{2+\beta-\epsilon}\right)^{\frac{1}{2-\beta+\epsilon}} \triangleq Q(n),$$

we will have

$$\Lambda(t) \le \frac{16\log(2n^3)}{n} \vee \frac{16\sqrt{M}\log(2n^3)}{\sqrt{n}}\overline{\rho}(t)^{\frac{\beta-\epsilon}{2}} \le \underline{\rho}(t),$$

which, according to Proposition 8 with $h=1$, we have $|T(t)-t| \le \frac{\Lambda(t)}{\underline{\rho}(t)} \le L(n)\frac{\Lambda(t)}{\rho(t)}$. Therefore, noticing that $\beta \le 1$, we have

$$\begin{aligned}
&\int_{\{t|\rho(t)\ge Q(n),|t|\le 2K\sqrt{\log n}\}} \rho(t)|T(t)-t|^2 dt \le L(n)^2 \int_{\{t|\rho(t)\ge Q(n),|t|\le 2K\sqrt{\log n}\}} \frac{\Lambda(t)^2}{\rho(t)} dt \\
&\le 4KL(n)^2\sqrt{\log n}\cdot \max_{\rho(t)\ge Q(n)} \frac{\Lambda(t)^2}{\rho(t)} \le 4KL(n)^2\sqrt{\log n}\cdot \max_{\rho(t)\ge Q(n)} \left\{\frac{256\log^2(2n^3)}{n^2\rho(t)} \vee \frac{256M\log^2(2n^3)}{n\rho(t)^{1+\epsilon-\beta}}\right\} \\
&\le 4KL(n)^2\sqrt{\log n}\cdot \left(\frac{256\log^2(2n^3)}{n^2Q(n)} \vee \frac{256M\log^2(2n^3)}{nQ(n)^{1+\epsilon-\beta}}\right) \\
&\le 4KL(n)^2\sqrt{\log n}\cdot \left(\frac{16\log(2n^3)}{nL(n)} \vee \left(\frac{256M\log^2(2n^3)}{n}\right)^{\frac{1}{2-\beta+\epsilon}} L(n)^{-\frac{(2+\beta-\epsilon)(1+\epsilon-\beta)}{2-\beta+\epsilon}}\right).
\end{aligned}$$

Further noticing that for any $\epsilon_1 > 0$, we have $L(n) = \mathcal{O}(n^{\epsilon_1})$. Hence for any $\epsilon' > 0$, we have

$$\int_{\{t|\rho(t)\ge Q(n),|t|\le 2K\sqrt{\log n}\}} \rho(t)|T(t)-t|^2 dt = \mathcal{O}\left(n^{-\frac{1}{2-\beta+\epsilon}+\epsilon'}\right).$$

As for those t with $\rho(t) < Q(n)$, according to (54) we have estimation

$$\begin{aligned}
&\int_{\{t|\rho(t)<Q(n),|t|\le 2K\sqrt{\log n}\}} \rho(t)|T(t)-t|^2 dt \le \int_{\{t|\rho(t)<Q(n),|t|\le 2K\sqrt{\log n}\}} \rho(t)(C_1+C_2\sqrt{\log n}|t|)^2 dx \\
&\le 4K\sqrt{\log n}\cdot Q(n)(C_1+2KC_2\log n)^2 = Q(n)\cdot\tilde{\mathcal{O}}(1) = \mathcal{O}\left(n^{-\frac{1}{2-\beta+\epsilon}+\epsilon'}\right)
\end{aligned}$$

Combine these two estimation together, we obtain that assuming event E, E_1, for any $\epsilon, \epsilon' > 0$, we have $\int_{|t|\le 2K\sqrt{n}} \rho(t)|T(t)-t|^2 dt = \mathcal{O}\left(n^{-\frac{1}{2-\beta+\epsilon}+\epsilon'}\right)$ and hence

$$\mathbb{E}\left[W_2^2(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})|E\cap E_1\right] = \int_{-\infty}^{\infty} \rho(t)|T(t)-t|^2 dt = \mathcal{O}\left(n^{-\frac{1}{2-\beta+\epsilon}+\epsilon'}\right) + \tilde{\mathcal{O}}(\frac{1}{n}) = \mathcal{O}\left(n^{-\frac{1}{2-\beta+\epsilon}+\epsilon'}\right).$$

Finally we notice that $\mathbf{P}[E^c \cup E_1^c] = \frac{C+1}{n^2}$, according to (53) we have

$$\mathbb{E}\left[W_2^2(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})\right] \le \mathbb{E}\left[W_2^2(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})|E\cap E_1\right] + \mathcal{O}\left(\frac{1}{n}\right) = \mathcal{O}\left(n^{-\frac{1}{2-\beta+\epsilon}+\epsilon'}\right).$$

Since ϵ and ϵ' can be chosen to be arbitrary small positive number, and $2\alpha = \frac{(1+K^2)^2}{2(1+K^4)} = \frac{1}{2-\beta}$, we have for any $\epsilon > 0$,

$$\mathbb{E}\left[W_2^2(\mathbb{P}*\mathcal{N},\mathbb{P}_n*\mathcal{N})\right] = \mathcal{O}\left(n^{-2\alpha+\epsilon}\right).$$

□

6 Proof of Theorem 3

Lemma 6. *Suppose $(X,Y)\sim\mathbb{P}_{X,Y}$, with marginal distributions $\mathbb{P}_X,\mathbb{P}_Y$. Let $\mathbb{P}_n$ be an empirical version of $\mathbb{P}_X$ generated with n samples. Then for every $1<\lambda\le 2$, we have*

$$\mathbb{E}[D_{KL}(\mathbb{P}_{Y|X}\circ\mathbb{P}_n\|\mathbb{P}_Y)]\le\frac{1}{\lambda-1}\log(1+\exp\{(\lambda-1)(I_\lambda(X;Y)-\log n)\})\,. \tag{57}$$

Proof. According to [VEH14], for any distribution $\mathbb{P},\mathbb{Q}$, the function $D_\lambda(\mathbb{P}\|\mathbb{Q})$ with respect to $\lambda\in(1,2]$ is non-decreasing, where D_λ is the Rényi divergence defined in Definition 1. Hence noticing from [VEH14] that for any distribution $\mathbb{P},\mathbb{Q}$, $\lim_{\lambda\to1}D_\lambda(\mathbb{P}\|\mathbb{Q})=D_{KL}(\mathbb{P}\|\mathbb{Q})$, we have

$$D_{KL}(\mathbb{P}_{Y|X}\circ\mathbb{P}_n\|\mathbb{P}_Y)\le D_\lambda(\mathbb{P}_{Y|X}\circ\mathbb{P}_n\|\mathbb{P}_Y).$$

Therefore, it is sufficient to prove that for any $1<\lambda\le2$,

$$\mathbb{E}[D_\lambda(\mathbb{P}_{Y|X}\circ\mathbb{P}_n\|\mathbb{P}_Y)]\le\frac{1}{\lambda-1}\log(1+\exp\{(\lambda-1)(I_\lambda(X;Y)-\log n)\}).$$

We suppose the n samples obtained in $\mathbb{P}_n$ to be $X_1,\cdots,X_n$, which satisfies that $(X_1,\cdots,X_n)\perp\!\!\!\perp Y$. According to the definition of Rényi divergence, Rényi mutual information and also the Jensen's inequality, we see that

$$\begin{aligned}\mathbb{E}[D_\lambda(\mathbb{P}_{Y|X}\circ\mathbb{P}_n\|\mathbb{P}_Y)]&=\frac{1}{\lambda-1}\mathbb{E}\left[\log\mathbb{E}\left[\left\{\frac{d(\mathbb{P}_{Y|X}\circ\mathbb{P}_n)(Y)}{d\mathbb{P}_Y(Y)}\right\}^\lambda\right]\middle|X_{1:n}\right]\\&\le\frac{1}{\lambda-1}\log\mathbb{E}\left[\left(\frac{d(\mathbb{P}_{Y|X}\circ\mathbb{P}_n)(Y)}{d\mathbb{P}_Y(Y)}\right)^\lambda\right].\end{aligned}\tag{58}$$

Then we introduced the channel $\mathbb{P}_{\bar Y|X_{1:n}}=\frac{1}{n}\sum_{i=1}^n\mathbb{P}_{Y|X=X_i}$ and we let $\mathbb{P}_{X_{1:n},\bar Y}=\mathbb{P}_{\bar Y|X_{1:n}}\circ\mathbb{P}_{X_{1:n}}$, where $\mathbb{P}_{X_{1:n}}=\mathbb{P}_X^{\otimes n}$ is the probability law of $X_{1:n}$. We notice that the marginal distribution of $\mathbb{P}_{X_{1:n},\bar Y}$ with respect to $\bar Y$ is exactly $\mathbb{P}_Y$. If we let $(X_{1:n},\bar Y)\sim\mathbb{P}_{X_{1:n}}\otimes\mathbb{P}_Y$, then we obtain that

$$\begin{aligned}I_\lambda(X_{1:n};\bar Y)&=\frac{1}{\lambda-1}\log\mathbb{E}\left[\left(\frac{d\mathbb{P}_{X_{1:n},\bar Y}(X_{1:n},Y)}{d\left[\mathbb{P}_{X_{1:n}}\otimes\mathbb{P}_Y(X_{1:n},Y)\right]}\right)^\lambda\right]\\&=\frac{1}{\lambda-1}\log\mathbb{E}\left[\left\{\frac{d\mathbb{P}_{Y|X_{1:n}}(Y|X_{1:n})}{d\mathbb{P}_Y(Y)}\right\}^\lambda\right]\\&=\frac{1}{\lambda-1}\log\mathbb{E}\left[\mathbb{E}\left[\left\{\frac{d(\mathbb{P}_{Y|X}\circ\mathbb{P}_n)(Y)}{d\mathbb{P}_Y(Y)}\right\}^\lambda\middle|X_{1:n}\right]\right]\\&=\frac{1}{\lambda-1}\log\mathbb{E}\left[\left(\frac{d(\mathbb{P}_{Y|X}\circ\mathbb{P}_n)(Y)}{d\mathbb{P}_Y(Y)}\right)^\lambda\right]\ge\mathbb{E}[D_\lambda(\mathbb{P}_{Y|X}\circ\mathbb{P}_n\|\mathbb{P}_Y)].\end{aligned}$$

Hence we only need to analyze $I_\lambda(X_{1:n};\bar Y)$. And we need to upper bound

$$\mathbb{E}\left[\left\{\frac{d\mathbb{P}_{Y|X_{1:n}}(Y|X_{1:n})}{d\mathbb{P}_Y(Y)}\right\}^\lambda\right]=\mathbb{E}\left[\left\{\frac{1}{n}\sum_{i=1}^n\frac{d\mathbb{P}_{Y|X}(Y|X_i)}{d\mathbb{P}_Y(Y)}\right\}^\lambda\right]. \tag{59}$$

Moreover, noticing that $(a+b)^{\lambda-1}\le a^{\lambda-1}+b^{\lambda-1}$ holds for $a,b>0$ and $1<\lambda\le2$, we have that for any n

$i.i.d.$ non-negative random variables B_i ($1 \le i \le n$),

$$\mathbb{E}\left[B_i\left(B_i+\sum_{j\neq i}B_j\right)^{\lambda-1}\right] \le \mathbb{E}[B_i\cdot B_i^{\lambda-1}] + \mathbb{E}\left[B_i\cdot\left(\sum_{j\neq i}B_j\right)^{\lambda-1}\right]$$
$$= \mathbb{E}[B_1^\lambda] + \mathbb{E}[B_i]\cdot\mathbb{E}\left[\left(\sum_{j\neq i}B_j\right)^{\lambda-1}\right]$$
$$\le \mathbb{E}[B_1^\lambda] + \mathbb{E}[B_1]\cdot\left(\sum_{j\neq i}\mathbb{E}[B_j]\right)^{\lambda-1} = \mathbb{E}[B_1^\lambda] + \mathbb{E}[B_1]\cdot((n-1)\mathbb{E}[B_1])^{\lambda-1},$$

where in the second inequality we use the Jensen's inequality. Therefore, summing up the above inequality for $1 \le i \le n$, we have

$$\mathbb{E}\left[\left\{\sum_{i=1}^n B_i\right\}^\lambda\right] \le n\mathbb{E}[B_1^\lambda] + n\cdot(n-1)^{\lambda-1}\left(\mathbb{E}[B_1]\right)^\lambda \le n\mathbb{E}[B_1^\lambda] + n^\lambda\left(\mathbb{E}[B_1]\right)^\lambda .$$

This puts us into a well-known setting of Rosenthal-type inequalities, which is known to be essentially tight [Sch11].

Next, since $Y \perp\!\!\!\perp (X_1, \cdots, X_n)$, for every fixed Y, random variables $\frac{d\mathbb{P}_{Y|X}(Y|X_i)}{d\mathbb{P}_Y(Y)}$ are $i.i.d.$ Hence choosing $B_i = \frac{d\mathbb{P}_{Y|X}(Y|X_i)}{d\mathbb{P}_Y(Y)}$, we obtain that

$$\mathbb{E}\left[\left\{\frac{1}{n}\sum_i\frac{d\mathbb{P}_{Y|X}(Y|X_i)}{d\mathbb{P}_Y(Y)}\right\}^\lambda\middle|Y\right] \le n^{-\lambda}\cdot\mathbb{E}\left[\left\{\sum_i\frac{d\mathbb{P}_{Y|X}(Y|X_i)}{d\mathbb{P}_Y(Y)}\right\}^\lambda\middle|Y\right]$$
$$\le n^{-\lambda}\cdot\left(n\cdot\mathbb{E}\left[\left\{\frac{d\mathbb{P}_{Y|X}(Y|X)}{d\mathbb{P}_Y(Y)}\right\}^\lambda\middle|Y\right] + n^\lambda\cdot\left(\mathbb{E}\left[\frac{d\mathbb{P}_{Y|X}(Y|X)}{d\mathbb{P}_Y(Y)}\middle|Y\right]\right)^\lambda\right)$$
$$\le n^{1-\lambda}\mathbb{E}\left[\left\{\frac{d\mathbb{P}_{Y|X}(Y|X)}{d\mathbb{P}_Y(Y)}\right\}^\lambda\middle|Y\right] + \left(\mathbb{E}\left[\frac{d\mathbb{P}_{Y|X}(Y|X)}{d\mathbb{P}_Y(Y)}\middle|Y\right]\right)^\lambda.$$

Using the fact that $X \perp\!\!\!\perp Y$ and hence $\mathbb{E}[\mathbb{P}_{Y|X}(Y|X)|Y] = \int_X P_{Y|X}(Y|X)d\mathbb{P}_X(X) = \int_X d\mathbb{P}_{X,Y}(X,Y) = \mathbb{P}_Y(Y)$, we notice that for any given Y,

$$\mathbb{E}\left[\frac{d\mathbb{P}_{Y|X}(Y|X)}{d\mathbb{P}_Y(Y)}\middle|Y\right] = \left.\frac{d\mathbb{E}[\mathbb{P}_{Y|X}(Y|X)]}{d\mathbb{P}_Y(Y)}\right|_Y = \left.\frac{d\mathbb{P}_Y(Y)}{d\mathbb{P}_Y(Y)}\right|_Y = 1.$$

Therefore, we can upper bound (59) as

$$\mathbb{E}\left[\left\{\frac{1}{n}\sum_{i=1}^n\frac{d\mathbb{P}_{Y|X}(Y|X_i)}{d\mathbb{P}_Y(Y)}\right\}^\lambda\right] = \mathbb{E}\left[\mathbb{E}\left[\left\{\frac{1}{n}\sum_{i=1}^n\frac{d\mathbb{P}_{Y|X}(Y|X_i)}{d\mathbb{P}_Y(Y)}\right\}^\lambda\right]\middle|Y\right]$$
$$\le n^{1-\lambda}\mathbb{E}\left[\mathbb{E}\left[\left\{\frac{d\mathbb{P}_{Y|X}(Y|X)}{d\mathbb{P}_Y(Y)}\right\}^\lambda\middle|Y\right]\middle|Y\right] + \mathbb{E}\left[\left(\mathbb{E}\left[\frac{d\mathbb{P}_{Y|X}(Y|X)}{d\mathbb{P}_Y(Y)}\middle|Y\right]\right)^\lambda\middle|Y\right]$$
$$\le n^{1-\lambda}\mathbb{E}\left[\left\{\frac{d\mathbb{P}_{Y|X}(Y|X)}{d\mathbb{P}_Y(Y)}\right\}^\lambda\right] + 1$$
$$= n^{1-\lambda}\cdot\exp\left((\lambda-1)I_\lambda(X;Y)\right) + 1.$$

This implies that

$$I_\lambda(X_{1:n};\bar{Y}) \le \frac{1}{\lambda-1}\log\left(1+n^{1-\lambda}\exp\{(\lambda-1)I_\lambda(X;Y)\}\right),$$

which together with (58) recovers (57).

$\square$

Remark 7. *Hayashi [Hay06] upper bounds the LHS of (57) with*

$$\frac{\lambda}{\lambda-1}\log\left(1+\exp\left\{\frac{\lambda-1}{\lambda}(K_\lambda(X;Y)-\log n)\right\}\right),$$

where $K_\lambda(X;Y)=\inf_{\mathbb{Q}_Y} D_\lambda(\mathbb{P}_{X,Y}\|\mathbb{P}_X\otimes\mathbb{Q}_Y)$ is the so-called Sibson-Csiszar information, cf. [Sib69]. This bound, however, does not have the right rate of convergence as $n\to\infty$, at least for $\lambda=2$ as comparison with Prop. 5 in [GGNWP20]. We note that [Hay06, HV93] also contain bounds on $\mathbb{E}[\mathrm{TV}(\mathbb{P}_{Y|X}\circ\mathbb{P}_n,\mathbb{P}_Y)]$ which do not assume existence of $\lambda>1$ moment of $\frac{\mathbb{P}_{Y|X}}{\mathbb{P}_Y}$ and instead rely on the distribution of $\log\frac{d\mathbb{P}_{Y|X}}{d\mathbb{P}_Y}$.

We are now ready to prove Theorem 3.

Proof of Theorem 3. We consider $X\sim\mathbb{P}, Z\sim\mathcal{N}(0,\sigma^2 I_d), X\perp\!\!\!\perp Z$ and $Y=X+Z$. Then conditioned on X, we have $Y\sim\mathcal{N}(X,\sigma^2 I_d)$, which indicates that $\mathbb{P}_{Y|X}\circ\mathbb{P}_n\sim\mathbb{P}_n*\mathcal{N}(0,\sigma^2 I_d)$. Therefore, adopting Lemma 6 and Lemma 1, we obtain that for any $1<\lambda<2$,

$$\begin{aligned}
&\mathbb{E}[D_{KL}(\mathbb{P}_n*\mathcal{N}(0,\sigma^2 I_d)\|\mathbb{P}*\mathcal{N}(0,\sigma^2 I_d))]\\
\le&\frac{1}{\lambda-1}\log(1+\exp((\lambda-1)(I_\lambda(X;Y)-\log n)))\\
\le&\frac{1}{\lambda-1}\cdot\exp((\lambda-1)(I_\lambda(X;Y)-\log n))\\
\le&\frac{C}{(\lambda-1)n^{\lambda-1}(2-\lambda)^d}.
\end{aligned}$$

Choosing $\lambda=2-\frac{1}{\log n}$, and noticing that

$$n^{\lambda-1}=n^{-\frac{1}{\log n}+1}=x\cdot\exp\left(-\log n\cdot\frac{1}{\log n}\right)=\frac{n}{e},$$

we have

$$\mathbb{E}[D_{KL}(\mathbb{P}_n*\mathcal{N}(0,\sigma^2 I_d)\|\mathbb{P}*\mathcal{N}(0,\sigma^2 I_d))]\le\frac{Ce(\log n)^d}{(1-1/\log n)n}=\mathcal{O}\left(\frac{(\log n)^d}{n}\right).$$

Hence (47) holds, which implies the upper bound part of Theorem 2. $\square$

Acknowledgments

This work was supported by the MIT-IBM Watson AI Lab. AB acknowledges support from the National Science Foundation Graduate Research Fellowship under Grant No. 1122374. AR acknowledges support from the Simons Foundation and the NSF through awards DMS-2031883 and DMS-1953181.

References

[Ber41] Andrew C Berry. The accuracy of the gaussian approximation to the sum of independent variates. *Transactions of the american mathematical society*, 49(1):122–136, 1941.

[BLG14] Emmanuel Boissard and Thibaut Le Gouic. On the mean speed of convergence of empirical and occupation measures in wasserstein distance. In *Annales de l'IHP Probabilités et statistiques*, volume 50, pages 539–563, 2014.

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[CCNW21] Hong-Bin Chen, Sinho Chewi, and Jonathan Niles-Weed. Dimension-free log-sobolev inequalities for mixture distributions. *Journal of Functional Analysis*, 281(11):109236, 2021.

[DSS13] Steffen Dereich, Michael Scheutzow, and Reik Schottstedt. Constructive quantization: Approximation by empirical measures. In *Annales de l'IHP Probabilités et statistiques*, volume 49, pages 1183–1203, 2013.

[Dud69] Richard Mansfield Dudley. The speed of mean glivenko-cantelli convergence. *The Annals of Mathematical Statistics*, 40(1):40–50, 1969.

[Dur19] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[Ess56] Carl-Gustav Esseen. A moment inequality with an application to the central limit theorem. *Scandinavian Actuarial Journal*, 1956(2):160–170, 1956.

[FG15] Nicolas Fournier and Arnaud Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3):707–738, 2015.

[GGNWP20] Ziv Goldfeld, Kristjan Greenewald, Jonathan Niles-Weed, and Yury Polyanskiy. Convergence of smoothed empirical measures with applications to entropy estimation. *IEEE Transactions on Information Theory*, 66(7):4368–4391, 2020.

[GK+06] Evarist Giné, Vladimir Koltchinskii, et al. Concentration inequalities and asymptotic results for ratio type empirical processes. *The Annals of Probability*, 34(3):1143–1216, 2006.

[Hay06] Masahito Hayashi. General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel. *IEEE Transactions on Information Theory*, 52(4):1562–1575, 2006.

[HV93] Te Sun Han and Sergio Verdú. Approximation theory of output statistics. *IEEE Transactions on Information Theory*, 39(3):752–772, 1993.

[PW16] Yury Polyanskiy and Yihong Wu. Wasserstein continuity of entropy and outer bounds for interference channels. *IEEE Transactions on Information Theory*, 62(7):3992–4002, 2016.

[Rén61] Alfréd Rényi. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 547–561. University of California Press, 1961.

[Sch11] Gideon Schechtman. Extremal configurations for moments of sums of independent positive random variables. In *Banach Spaces and their Applications in Analysis*, pages 183–192. De Gruyter, 2011.

[Sib69] Robin Sibson. Information radius. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 14(2):149–160, 1969.

[VEH14] Tim Van Erven and Peter Harremos. Rényi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014.

[Ver18] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[Vil03] Cédric Villani. *Topics in optimal transportation.* Number 58. American Mathematical Soc., 2003.

[WB19] Jonathan Weed and Francis Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in wasserstein distance. *Bernoulli*, 25(4A):2620–2648, 2019.

[WW+16] Feng-Yu Wang, Jian Wang, et al. Functional inequalities for convolution probability measures. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 52, pages 898–914. Institut Henri Poincaré, 2016.

[Zim13] David Zimmermann. Logarithmic sobolev inequalities for mollified compactly supported measures. *Journal of Functional Analysis*, 265(6):1064–1083, 2013.

A Proof of Subgaussianity in Section 3

Proposition 11. *Given positive constant $c > 2, c_1 > 0$, we consider the distribution $\mathbb{P} = \sum_{k=0}^{\infty} p_k \delta_{r_k}$, with $r_0 = 0, r_1 = 1, r_{i+1} = cr_i, \forall i \geq 1$, and also the*

$$
\begin{aligned}
p_k &= c_1 \exp\left(-\frac{r_k^2}{2K^2}\right), && k \geq 1,\\
p_k &= 1 - \sum_{k=1}^{\infty} p_k, && k = 0.
\end{aligned}
$$

Then there exists some $c_1 > 0$ such that for any constant $c > 2$, we have $c_1 \cdot \sum_{k=1}^{\infty} \exp\left(-\frac{r_k^2}{2K^2}\right) < 1$, and also distribution $\mathbb{P}$ is a K-SubGaussian distribution, i.e. for $S \sim \mathbb{P}$,

$$
\mathbb{E}\left[\exp\left(\alpha\left(S - \mathbb{E}[S]\right)\right)\right] \leq \exp\left(\frac{K^2\alpha^2}{2}\right), \quad \forall \alpha \in \mathbb{R}.
$$

Proof. We let

$$
S_1 = \mathbb{E}[S] = \sum_{k=0}^{\infty} k p_k \geq 0.
$$

(Here S_1 is only a real number, not a random variable.) Then we have

$$
\sum_{k=1}^{\infty} p_k \leq c_1 \sum_{k=1}^{\infty} \exp\left(-\frac{k}{2K^2}\right) \leq \frac{c_1}{1 - \exp\left(-\frac{1}{2K^2}\right)}
$$

and also

$$
S_1 = c_1 \sum_{k=1}^{\infty} k \exp\left(-\frac{r_k^2}{2K^2}\right) \leq c_1 \sum_{k=1}^{\infty} k \exp\left(-\frac{k}{2K^2}\right) = c_1 \cdot \frac{\exp\left(-\frac{1}{2K^2}\right)}{\left(1 - \exp\left(-\frac{1}{2K^2}\right)\right)^2}.
$$

We will choose c_1 close to 0 enough such that $\sum_{k=1}^{\infty} p_k \leq \frac{1}{2}$ hence $p_0 = 1 - \sum_{k=1}^{\infty} \geq 0$ is well defined. In order to prove the subgaussian property, we define

$$
\begin{aligned}
f(\alpha) &\triangleq \exp\left(-\frac{K^2\alpha^2}{2}\right) \cdot \mathbb{E}\left[\exp(\alpha(S - S_1))\right]\\
&= \exp\left(-\frac{K^2\alpha^2}{2} - \alpha S_1\right) \cdot \left(p_0 + \sum_{k=1}^{\infty} p_k \exp(\alpha r_k)\right)\\
&= \exp\left(-\frac{K^2\alpha^2}{2} - \alpha S_1\right) \cdot \left(p_0 + c_1 \sum_{k=1}^{\infty} \exp\left(-\frac{r_k^2}{2K^2} + \alpha r_k\right)\right)\\
&= \exp\left(-\frac{K^2\alpha^2}{2} - \alpha S_1\right) \cdot \left(p_0 + c_1 \sum_{k=1}^{\infty} \exp\left(-\frac{1}{2K^2}\left(r_k - \alpha K^2\right)^2\right) \exp\left(\frac{K^2\alpha^2}{2}\right)\right)\\
&= p_0 \exp\left(-\frac{K^2\alpha^2}{2} - \alpha S_1\right) + c_1 \sum_{k=1}^{\infty} \exp\left(-\frac{1}{2K^2}\left(r_k - \alpha K^2\right)^2 - \alpha S_1\right).
\end{aligned}
$$

To prove that $f(\alpha) \leq 1$ for every $\alpha \in \mathbb{R}$, we consider cases where $\alpha K^2 \geq \frac{1}{4}$ and $\alpha K^2 \leq -2S_1$ and $-1 \leq \alpha K^2 < \frac{1}{4}$ respectively (if we can choose c_1 such that $2S_1 \leq 1$ holds for every c, then these three cases cover all the situations).

1. When $\alpha K^2 \le -2S_1$, we have

$$
\begin{aligned}
f(\alpha) &= p_0 \exp\left(-\frac{K^2\alpha^2}{2} - \alpha S_1\right) + c_1 \sum_{k=1}^{\infty} \exp\left(-\frac{1}{2K^2}\left(r_k - \alpha K^2\right)^2 - \alpha S_1\right) \\
&\le p_0 \exp\left(-\frac{K^2\alpha^2}{2} - \alpha S_1\right) + c_1 \sum_{k=1}^{\infty} \exp\left(-\frac{r_k^2 + \alpha^2 K^4}{2K^2} - \alpha S_1\right) \\
&= \left(p_0 + \sum_{k=1}^{\infty} p_k\right) \cdot \exp\left(-\frac{K^2\alpha^2}{2} - \alpha S_1\right) \\
&\le \exp\left(-\frac{K^2\alpha^2}{2} - \alpha S_1\right) \le 1.
\end{aligned}
$$

2. When $\alpha K^2 \ge \frac{1}{4}$, we have

$$
p_0 \exp\left(-\frac{K^2\alpha^2}{2} - \alpha S_1\right) \le p_0 \exp\left(-\frac{1}{8K^2}\right) \le \exp\left(-\frac{1}{8K^2}\right)
$$

Moreover, we suppose k_0 to be the smallest k such that $r_k - \alpha K^2$ to be positive. Since $r_{k+1} - r_k \ge 1$ for every k, we have for $k \ge k_0$, $r_k - \alpha K^2 \ge k - k_0 + r_{k_0} - \alpha K^2 \ge k - k_0$, and for $k < k_0$, $r_k - \alpha K \le r_{k_0-1} - \alpha K + (k_0 - 1 - k) \le k_0 - 1 - k$ since $r_{k_0-1} \le 0$. Therefore, we have

$$
\begin{aligned}
&\sum_{k=1}^{\infty} \exp\left(-\frac{1}{2K^2}\left(r_k - \alpha K^2\right)^2 - \alpha S_1\right) \\
\le& \sum_{k=1}^{\infty} \exp\left(-\frac{1}{2K^2}\left(r_k - \alpha K^2\right)^2\right) \\
=& \sum_{k=1}^{k_0-1} \exp\left(-\frac{(r_k - \alpha K^2)^2}{2K^2}\right) + \sum_{k=k_0}^{\infty} \exp\left(-\frac{(r_k - \alpha K^2)^2}{2K^2}\right) \\
\le& \sum_{k=1}^{k_0-1} \exp\left(-\frac{k_0 - 1 - k}{2K^2}\right) + \sum_{k=k_0}^{\infty} \exp\left(-\frac{k - k_0}{2K^2}\right) \\
\le& \sum_{k=0}^{\infty} \exp\left(-\frac{1}{2K^2}\right)^k + \sum_{k=0}^{\infty} \exp\left(-\frac{1}{2K^2}\right)^k \\
=& \frac{2}{1 - \exp\left(-\frac{1}{2K^2}\right)}.
\end{aligned}
$$

Hence if

$$
c_1 \le \frac{1}{2}\left(1 - \exp\left(-\frac{1}{8K^2}\right)\right)\left(1 - \exp\left(-\frac{1}{2K^2}\right)\right),
$$

we would have

$$
p_0 \exp\left(-\frac{K^2\alpha^2}{2}\right) + c_1 \sum_{k=1}^{\infty} \exp\left(-\frac{1}{2K^2}\left(r_k - \alpha K\right)^2\right) \le \exp\left(-\frac{1}{8K^2}\right) + c_1 \cdot \frac{2}{1 - \exp\left(-\frac{1}{2K^2}\right)} \le 1.
$$

3. When $-1 \le \alpha K^2 < \frac{1}{4}$, we calculate that

$$
h(\alpha) \triangleq \exp\left(\frac{K^2\alpha^2}{2} + \alpha S_1\right) \cdot f'(\alpha) = -p_0(\alpha K^2 + S_1) + c_1 \sum_{k=1}^{\infty} \left(r_k - \alpha K^2 - S_1\right) \exp\left(-\frac{r_k^2}{2K^2} + \alpha r_k\right)
$$

and

$$\begin{aligned}
h'(\alpha) &= -p_0K^2 + c_1\sum_{k=1}^{\infty}\left(r_k^2 - \alpha K^2 r_k - S_1 r_k - K^2\right)\exp\left(-\frac{r_k^2}{2K^2} + \alpha r_k\right)\\
&\le -p_0K^2 + c_1\sum_{k=1}^{\infty}\left(r_k^2 - \alpha K^2 r_k\right)\exp\left(-\frac{r_k^2}{2K^2} + \alpha r_k\right)\\
&\le -p_0K^2 + c_1\sum_{k=1}^{\infty}\left(r_k^2 - \alpha K^2 r_k\right)\exp\left(-\frac{r_k^2}{2K^2} + \frac{r_k}{4K^2}\right)\\
&\le -p_0K^2 + 2c_1\sum_{k=1}^{\infty}r_k^2\exp\left(-\frac{r_k^2}{4K^2}\right),
\end{aligned}$$

where we use the fact that $r_k \ge 1$ for any $k \ge 1$. We then notice that function $g(x) = x^2\exp\left(-\frac{x^2}{4K^2}\right)$ is monotonically decreasing when $x \ge 2K$. Hence for $k \ge 2K+1$ we have $r_k \ge 2K+1$ and

$$\begin{aligned}
&\sum_{k\ge 2K+1}^{\infty} r_k^2\exp\left(-\frac{r_k^2}{4K^2}\right)\\
&\le \int_{2K}^{\infty} x^2\exp\left(-\frac{x^2}{4K^2}\right)dx \le 3K^3.
\end{aligned}$$

For those $k < 2K+1$, there are at most $2K+1$ number of such K, and for each of such k we have

$$r_k^2\exp\left(-\frac{r_k^2}{4K^2}\right) = K^2\cdot\left(\frac{r_k}{K}\right)^2\exp\left(-\frac{1}{4}\left(\frac{r_k}{K}\right)^2\right) \le 2K^2.$$

Therefore, we have

$$\sum_{k=1}^{\infty} r_k^2\exp\left(-\frac{r_k^2}{4K^2}\right) \le 3K^3 + (2K+1)K^2 \le 6K^3.$$

Hence when $c_1 < \frac{1}{24}$ and $p_0 \ge \frac{1}{2}$, we have $h'(\alpha) \le 0$ for every $-1 \le \alpha K^2 \le \frac{1}{4}$. Moreover, we can calculate that

$$h(0) = p_0S_1 + c_1\sum_{k=1}^{\infty}(r_k - S_1)\exp\left(-\frac{r_k^2}{2K^2}\right) = p_0S_1 + \sum_{k=1}^{\infty}p_k(r_k - S_1) = \mathbb{E}[S] - S_1 = 0.$$

This indicates that for $-1/K^2 \le \alpha \le 0$, we have $h(\alpha) \ge 0$ hence $f'(\alpha) \ge 0$, and for $0 \le \alpha \le 1/(4K^2)$, we have $h(\alpha) \le 0$ hence $f'(\alpha) \le 0$. This leads to

$$f(\alpha) \le f(0) = p_0 + c_1\sum_{k=1}^{\infty}\exp\left(-\frac{r_k^2}{2K^2}\right) = \sum_{k=0}^{\infty}p_k = 1$$

holds for every $-1/K^2 \le \alpha \le 1/(4K^2)$.

Above all, if we choose c_1 such that the following items hold, then we will have $f(\alpha) \le 1$ for all $\alpha \in \mathbb{R}$:

1. $2S_1 \le 1$, which can be obtained from $c_1 \le \frac{\left(1-\exp\left(-\frac{1}{2K^2}\right)\right)^2}{2\exp\left(-\frac{1}{2K^2}\right)}$;
2. $c_1 \le \frac{1}{24}$;
3. $c_1 \le \frac{1}{2}\left(1 - \exp\left(-\frac{1}{8K^2}\right)\right)\left(1 - \exp\left(-\frac{1}{2K^2}\right)\right)$;

4. $1-p_0=\sum_{k=1}^{\infty} p_k \le \frac{1}{2}$, which can be obtained from $c_1 \le \frac{1-\exp\left(-\frac{1}{2K^2}\right)}{2}$.

Hence if we choose

$$c_1=\min\left\{\frac{1}{24}, \frac{\left(1-\exp\left(-\frac{1}{2K^2}\right)\right)^2}{2\exp\left(-\frac{1}{2K^2}\right)}, \frac{1}{2}\left(1-\exp\left(-\frac{1}{8K^2}\right)\right)\left(1-\exp\left(-\frac{1}{2K^2}\right)\right), \frac{1-\exp\left(-\frac{1}{2K^2}\right)}{2}\right\},$$

and p_k in (22), we would have $f(\alpha)\le 1$ for all $\alpha\in\mathbb{R}$. Therefore, we have

$$\mathbb{E}\left[\exp(\alpha(S-S_1))\right] \le \exp\left(\frac{K^2\alpha^2}{2}\right), \quad \forall \alpha\in\mathbb{R},$$

which indicates that distribution P is a K-subgaussian. □

B LSI and T_2 constants for Bernoulli-Gaussian mixtures

B.1 Proof of the Non-Existence of Uniform Bound of LSI Constants for Bernoulli Distributions in 4.1

In this subsection, we will prove that for the Bernoulli distribution class in Section 4.1, there constants in the corresponding log-Sobolev inequalities do not have a uniform bound.

Theorem 4. *Suppose σ is a given constant which is smaller than K. Consider the following Bernoulli distributions:*

$$\mathbb{P}_h=(1-p_h)\delta_0+p_h\delta_h, \quad p_h=\exp\left(-\frac{h^2}{2K^2}\right).$$

We use C_h to denote the constant of LSI of distribution $\mu_h=\mathbb{P}_h * \mathcal{N}(0,\sigma^2)$: C_h is the smallest constant such that for any smoothed, compact supported function f such that $\int_{\mathbb{R}} f^2 d\mu_h=1$, we have

$$\int_{\mathbb{R}} f^2\log f^2 d\mu_h \le C_h\int_{\mathbb{R}} |f'|^2 d\mu.$$

Then we have

$$\sup_{h\in\mathbb{R}_+} C_h=\infty.$$

Proof of Theorem 4. We choose $x_1<-1<0<x_2<h-1$, where x_1 and x_2 are determined later, and we let

$$f_h(x)=\begin{cases}0 & x\le x_1,\\ t(x-x_1) & x_1\le x\le x_1+1,\\ t & x_1+1\le x\le x_2,\\ -t(x-x_2-1) & x\ge x_2,\end{cases}$$

where t is the constant chosen such that $\int_{\mathbb{R}} f_h^2 d\mu_h=1$. Then f_h is a continuous function on $\mathbb{R}$, and $|f_h'(x)|\le t$ for any $x\in\mathbb{R}$. (Notice here f_h is not a smooth function, but it has only finite points which are not smoothed. Hence after some simple smoothing procedure near these points, e.g. convolved with some mollifier, we can construct a sequence of functions converging to f_h such that if the LSI works for functions in this sequence, the LSI also works for f_h.) Next, we will calculate the lower bound of C_h such that the LSI works for function f_h. We denote

$$q_{h,1}=\mu_h((-\infty,x_1]), \quad q_{h,2}=\mu_h((x_1,x_1+1]), \quad q_{h,3}=\mu_h((x_1+1,x_2]),$$
$$q_{h,4}=\mu_h((x_2,x_2+1]), \quad q_{h,5}=\mu_h((x_2+1,\infty)).$$

Then we have

$$q_{h,1}+q_{h,2}+q_{h,3}+q_{h,4}+q_{h,5}=1.$$

According to the definition of f, we have

$$1=\int_{\mathbb{R}} f_h^2 d\mu_h \le (q_{h,2}+q_{h,3}+q_{h,4})t^2,$$

which indicates that $t^2 \ge \frac{1}{q_{h,2}+q_{h,3}+q_{h,4}} \ge 1$. Since for any $a \ge 0$, we have $a\log a \ge -1$, we also have

$$\int_{\mathbb{R}} f_h^2 \log f_h^2 d\mu_h \ge q_{h,3}t^2\log t^2 - (q_{h,2}+q_{h,4}) \ge f_h^2 d\mu_h \ge q_{h,3}t^2\log t^2-(q_{h,2}+q_{h,4})t^2.$$

Moreover, we also notice that $|f_h'(x)|^2=t^2$ if $x\in(x_1,x_1+1)\cup(x_2,x_2+1)$, while $|f_h'(x)|^2=0$ for other x. Therefore, we obtain that

$$\int_{\mathbb{R}} |f_h'|^2 d\mu_h = (q_{h,2}+q_{h,4})t^2.$$

Hence if we require the LSI with constant C_h holds for f_h, we will have

$$q_{h,3}t^2\log t^2-(q_{h,2}+q_{h,4})t^2 \le C_h(q_{h,2}+q_{h,4})t^2,$$

which indicates that

$$\begin{aligned} C_h &\ge \frac{q_{h,3}\log t^2}{q_{h,2}+q_{h,4}}-1 \ge \frac{-q_{h,3}\log(q_{h,2}+q_{h,3}+q_{h,4})}{q_{h,2}+q_{h,4}}-1 \\ &= \frac{-q_{h,3}\log(1-q_{h,1}-q_{h,5})}{q_{h,2}+q_{h,4}}-1 \ge \frac{q_{h,3}(q_{h,1}+q_{h,5})}{q_{h,2}+q_{h,4}}-1 \ge \frac{q_{h,3}q_{h,5}}{q_{h,2}+q_{h,4}}-1. \end{aligned}$$

We use $\varphi_{\sigma^2}(x)$ to denote the PDF of $\mathcal{N}(0,\sigma^2)$ at point x. According to the definition of μ_h, and also noticing that $0<x_1<h-1$, we have

$$q_{h,4}=\int_{x_1}^{x_1+1}(1-p_h)\varphi_{\sigma^2}(x)+p_h\varphi_{\sigma^2}(x-h)dx \le \varphi_{\sigma^2}(x)+p_h\varphi_{\sigma^2}(h-x-1),$$

and also

$$q_{h,5}=\int_{x_1+1}^{\infty}(1-p_h)\varphi_{\sigma^2}(x)+p_h\varphi_{\sigma^2}(x-h)dx \ge \int_{x_1+1}^{\infty}p_h\varphi_{\sigma^2}(x-h)dx \ge \int_h^{\infty}p_h\varphi_{\sigma^2}(x-h)dx=\frac{p_h}{2}.$$

We further notice that $\lim_{x_1\to-\infty}q_{h,1}=\lim_{x_1\to-\infty}q_{h,2}=0$. Hence letting $x_1\to-\infty$, we will obtain that C_h satisfies

$$C_h \ge \lim_{x_1\to-\infty}\frac{q_{h,3}q_{h,5}}{q_{h,2}+q_{h,4}}-1=\lim_{x_1\to-\infty}\frac{q_3q_5}{q_4}-1=\frac{(1-q_4-q_5)q_5}{q_4}-1 \ge \frac{(1-q_5)q_5}{q_4}-2.$$

When $\sigma<K$, we will choose $x=h\sqrt{\sigma/K}$, then we will have $\lim_{h\to\infty}x-h-1=\infty$, which indicates that

$$0\le\lim_{h\to\infty}\frac{q_{h,4}}{p_h}=\lim_{h\to\infty}\frac{\varphi_{\sigma^2}(h\sqrt{\sigma/K})+p_h\exp\varphi(h(1-\sqrt{\sigma/K}))}{p_h}=0,$$

and also

$$0\le\lim_{h\to\infty}q_{h,5}\le\lim_{h\to\infty}\int_{h\sqrt{\sigma/K}+1}^{\infty}\varphi_{\sigma^2}(x)dx+\lim_{h\to\infty}p_h=0,$$

which indicates that $\lim_{h\to\infty}(1-q_{h,5})=1$. Above all, we obtain that

$$\lim_{h\to\infty}\frac{(1-q_5)q_5}{q_4}-2=\infty,$$

which indicates that $\lim_{h\to\infty}C_h=\infty$, and the uniform bound for C_h does not exists. $\square$

B.2 Proof of the Transportation-Entropy Inequality Constant

Theorem 5. *Suppose σ is a given constant which is smaller than K. Consider the following Bernoulli distributions:*

$$\mathbb{P}_h = (1-p_h)\delta_0 + p_h\delta_h, \quad p_h = \exp\left(-\frac{h^2}{2K^2}\right).$$

We use C'_h to denote the constant of transportation-entropy inequality : C_h is the smallest constant such that

$$W_2(\mathbb{P}_h * \mathcal{N}(0,\sigma^2), \mathbb{Q}) \le C'_h D_{KL}(\mathbb{P}_h * \mathcal{N}(0,\sigma^2) \| \mathbb{Q}) \qquad \forall \text{ distribution } \mathbb{Q}. \tag{60}$$

Then we have

$$\sup_{h\in\mathbb{R}_+} C'_h = \infty.$$

Proof. We let $\mathbb{Q}_h = (1-q_h)\delta_0 + q_h\delta_h$ with $q_h = p_h - \exp\left(-\frac{(1-\delta)(1+\sigma^2/K^2)^2h^2}{8\sigma^2}\right)$ for some δ smaller enough such that $(1-\delta)(1+\sigma^2/K^2)^2h^2 > 4\sigma^2/K^2$. According to data-processing inequality we have

$$\begin{aligned}
&D_{KL}(\mathbb{P}_h * \mathcal{N}(0,\sigma^2)\|\mathbb{Q}_h * \mathcal{N}(0,\sigma^2)) \le D_{KL}(\mathbb{P}_h\|\mathbb{Q}_h) = p_h \log\frac{p_h}{q_h} + (1-p_h)\log\frac{1-p_h}{1-q_h} \\
&= -p_h\log\left(1+\frac{q_h-p_h}{p_h}\right) - (1-p_h)\log\left(1+\frac{p_h-q_h}{1-p_h}\right) \\
&\le -p_h\cdot\frac{q_h-p_h}{p_h} + p_h\cdot\frac{(q_h-p_h)^2}{p_h^2} - (1-p_h)\cdot\frac{p_h-q_h}{1-p_h} + (1-p_h)\cdot\frac{(q_h-p_h)^2}{(1-p_h)^2} \\
&\le 2\exp\left(\frac{h^2}{2K^2}\right)(p_h-q_h)^2,
\end{aligned}$$

where in the second inequality we use the fact that $-\log(1+x) \le -x + x^2$ for $x \ge -1/2$ and $\frac{q_h-p_h}{p_h} \ge -1/2$. Similar to the proof of Proposition 3, and noticing that $F_{q,h}(t) - F_{p,h}(t) = (q_h - p_h)(\Phi_\sigma(t) - \Phi_\sigma(t-h))$ where $F_{q,h}, F_{p,h}, \Phi_\sigma$ are CDFs of distribution $\mathbb{Q}_h * \mathcal{N}(0,\sigma^2), \mathbb{P}_h * \mathcal{N}(0,\sigma^2), \mathcal{N}(0,\sigma^2)$. We can prove that

$$W_2(\mathbb{P}_h * \mathcal{N}(0,\sigma^2), \mathbb{Q}_h * \mathcal{N}(0,\sigma^2))^2 = \Omega\left(\exp\left(-\frac{(1-\delta)(1+\sigma^2/K^2)^2h^2}{8\sigma^2}\right)\right)$$

while

$$D_{KL}(\mathbb{P}_h * \mathcal{N}(0,\sigma^2)\|\mathbb{Q}_h * \mathcal{N}(0,\sigma^2)) = \mathcal{O}\left(\frac{h^2}{2K^2} - \frac{(1-\delta)(1+\sigma^2/K^2)^2h^2}{4\sigma^2}\right).$$

Since $(1-\delta)(1+\sigma^2/K^2)^2h^2 > 4\sigma^2/K^2$, letting $h\to\infty$ we obtain that $\sup_{h\in\mathbb{R}_+} C'_h = \infty$. □