Non-parametric threshold for smoothed empirical Wasserstein distance

by

Zeyu Jia

B.S. in Information and Computing Science, Peking University (2020)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Computer Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2022

Author . . . .
Department of Electrical Engineering and Computer Science
December 13, 2021

Certified by . . . .
Yury Polyanskiy
Associate Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Certified by . . . .
Sasha Rakhlin
Professor of Statistics and Data Science Center
Thesis Supervisor

Accepted by . . . .
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Non-parametric threshold for smoothed empirical Wasserstein distance

by

Zeyu Jia

Submitted to the Department of Electrical Engineering and Computer Science
on December 13, 2021, in partial fulfillment of the
requirements for the degree of
Master of Science in Computer Science and Engineering

Abstract

Consider an empirical measure $\mathbb{P}_n$ induced by n iid samples from a d-dimensional K-subgaussian distribution $\mathbb{P}$. We show that when $K < \sigma$, the Wasserstein distance $W_2^2(\mathbb{P}_n * \mathcal{N}(0, \sigma^2 I_d), \mathbb{P} * \mathcal{N}(0, \sigma^2 I_d))$ converges at the parametric rate $O(1/n)$, and when $K > \sigma$, there exists a K-subgaussian distribution $\mathbb{P}$ such that $W_2^2(\mathbb{P}_n * \mathcal{N}(0, \sigma^2 I_d), \mathbb{P} * \mathcal{N}(0, \sigma^2 I_d)) = \omega(1/n)$. This resolves the open problems in[7], closes the gap between where we get parametric rate and where we do not have parametric rate. Our result provides a complete characterization of the range of parametric rates for subgaussian P.

In addition, when $\sigma < K$, we establish more delicate results about the convergence rate of W2 distance squared. Assuming the distribution is one dimensional, we provide both the lower bound and the upper bound, demonstrating that the rate changes gradually from $\Theta(1/\sqrt{n})$ to $\Theta(1/n)$ as σ/K goes from 0 to 1. Moreover, we also establish that $D_{KL}(\mathbb{P}_n * \mathcal{N}(0, \sigma^2 I_d) \| \mathbb{P} * \mathcal{N}(0, \sigma^2 I_d)) = \tilde{O}(1/n)$. These results indicate a dichotomy of the convergence rate between the W2 distance squared and the KL divergence, resulting in the failure of T_2-transportation inequality when $\sigma < K$, hence also resolving the open problem in [17] about whether $K < \sigma$ is necessary in proving whether the log-Sobolev inequality holds for $\mathbb{P} * \mathcal{N}(0, \sigma^2)$.

Thesis Supervisor: Yury Polyanskiy
Title: Associate Professor of Electrical Engineering and Computer Science

Thesis Supervisor: Sasha Rakhlin
Title: Professor of Statistics and Data Science Center

Acknowledgments

First I would like to show my gratitude to my supervisors, Yury Polyanskiy and Sasha Rakhlin. Every time after discussing with my supervisors, I can always get enlightened. They showed me how to find meaningful problems in research, how to formulate into the form we can solve, and how to solve them in the most efficient way as well. They also provide many useful instructions on how to write papers in a proper way, which benefits my not only in the write-up of this thesis, and also among all the writings I will do in the future.

I would thank to Adam Block, who also provide some ideas and critical comments in this paper. I would also thank to Yuzhou Gu, Chenghao Guo and many other students in MIT. The discussion with them along my graduate study is the source of many ideas of mine.

I would like to thank to Yuan Shen and Yaodong Wang at Department of Electrical Engineering in Tsinghua University. During the 2020-2021 academic year, I cannot come to the MIT campus to pursue my graduate study, and I spent that time in visiting Prof. Yuan and Prof. Wang's group. Many ideas in this thesis was obtained during that part of time. I would also like to thank to students in Prof. Shen and Prof. Wang's group. They created a very great academic atmosphere so that I can carry my study more fluently during my visiting.

During the past 1 year and half, I was supported by 9 months of Seneff-Zue MIT EECS Fellowship, and 6 months of MIT-IBM funding. I would like to show great thanks to Seneff-Zue and also MIT-IBM for the support which makes my study and research possible at MIT.

I would show my greatest thanks to my wife Xiwen Hu. We married just a week before we come to America. Along my graduate study, she always offers her company and support. She always encourages me to keep going on or try to think through other ways when I meet difficulties in my research. This thesis would be impossible without her.

Contents

List of Figures

Chapter 1

Introduction

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 very large even if one restricts to convergence in Wasserstein W_p-distances, cf. [16, 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. Indeed, already in [4] 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 suprema over expectations of Lipschitz functions. However, his idea of recursive partitioning was cleverly adapted to the realm of couplings in [1], recovering Dudley's convergence rate of $n^{-1/d}$ also for $p > 1$. See [3, 5, 18] for more on this line of work, and also for a thorough survey of the recent literature.

We see that while $\mathbb{P}_n \to \mathbb{P}$ in W_p-distance, the convergence rate slows down as dimension d increases. Unfortunately, the rate of $n^{-1/d}$ is impractically slow already for moderate d. It turned out [7], however, 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. More precisely, let $\varphi_{\sigma^2 I_d}(x) \triangleq (2\pi\sigma^2)^{-d/2} e^{-\frac{\|x\|^2}{2\sigma^2}}$ be the density of $\mathcal{N}(0, \sigma^2 I_d)$ (if $d = 1$, we simply write $\varphi_{\sigma^2 I_d}(\cdot)$ as $\varphi_\sigma(\cdot)$), and for any probability measure $\mathbb{P}$ on $\mathbb{R}^d$ we define the convolved measure via

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

where E is any Borel set. Then [7, Prop. 6] shows

$$\mathbb{E}[W_2^2(\mathbb{P}_n * \mathcal{N}(0, \sigma^2 I_d), \mathbb{P} * \mathcal{N}(0, \sigma^2 I_d))] \leq \frac{C(d, \sigma, K)}{n}, \tag{1.1}$$

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

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

Note that in (1.1) constant C does not depend on $\mathbb{P}$. The (1.1) 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 [7, Corollary 2]: for $K \geq \sqrt{2}\sigma$ there exists a K-subgaussian distribution $\mathbb{P}$ in $\mathbb{R}^1$ such that

$$\lim_{n\to\infty} n\mathbb{E}[W_2^2(\mathbb{P}_n * \mathcal{N}(0, \sigma^2 I_d), \mathbb{P} * \mathcal{N}(0, \sigma^2 I_d))] = \infty, \tag{1.2}$$

where the expectation is with respect to n samples according to $\mathbb{P}$. We say that the rate of convergence is "parametric" if (1.1) holds and otherwise call it "non-parametric". Thus, the results of [7] shows that parametric rate for smoothed-W_2 is only attained by sufficiently light-tailed distributions $\mathbb{P}$ as measured by subgaussian constant.

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 (1.1), while

for $K > \sigma$ we have (1.2) 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 [7, 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

$$\rho = \rho(K, \sigma, d) \triangleq \lim_{n\to\infty} \sup_{\mathbb{P}-K\text{-subgaussian}} -\frac{\log \mathbb{E}[W_2(\mathbb{P}_n * \mathcal{N}(0, \sigma^2 I_d), \mathbb{P} * \mathcal{N}(0, \sigma^2 I_d))]}{\log n} \tag{1.3}$$

Previously, it was only known that $\frac{1}{4} \le \rho \le \frac{1}{2}$ for all $K > \sigma$ (note that (1.2) strongly suggests but does not formally imply $\rho < \frac{1}{2}$). Theorem 2 shows new upper and lower bounds for $d = 1$, which, albeit non-matching, demonstrate that $\rho \uparrow 1/2$ as $K \uparrow \sigma$ and $\rho \downarrow 1/4$ as $K \downarrow 0$.

3. So we can see that for a class of K-subgaussian distributions 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

$$\mathbb{E}[D(\mathbb{P}_n * \mathcal{N}(0, \sigma^2 I_d) \| \mathbb{P} * \mathcal{N}(0, \sigma^2 I_d))] \le \begin{cases} \frac{C(\sigma,K,d)\log^{d+1} n}{n}, & K > \sigma \\ \frac{C(\sigma,K,d)}{n}, & K < \sigma \end{cases}, \tag{1.4}$$

where $D(\mu\|\nu) = \int d\nu f(x) \log f(x)$, $f \triangleq \frac{d\mu}{d\nu}$ 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 non-parametric regime, the W_2 distance experiences a polynomial slow-down in rate, while KL only gets hit by (at most) poly-logarithmic penalty.

To better understand 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

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

The infimum over all such constants is denoted by $T_2(\nu)$. Talagrand originally demonstrated that $T_2(\varphi_\sigma) < \infty$. It turns out that $T_2(\mathbb{P} * \varphi_\sigma) < \infty$ as well. This was first shown by [19] for compactly supported $\mathbb{P}$ and extended to K-subgaussian $\mathbb{P}$ with $K < \sigma$ in [17] (in fact, both papers establish a stronger log-Sobolev inequality (LSI)).

Now comparing (1.4) 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 $\mathbb{P} * \varphi_\sigma$ does not satisfy T_2-transportation inequality (and hence does not satisfy the LSI either), that is $T_2(\mathbb{P} * \varphi_\sigma) = \infty$.*

We remark that it is straightforward to show that

$$\sup\{T_2(\mathbb{P} * \varphi_\sigma) : \mathbb{P} - K\text{-subgaussian}\} = \infty$$

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 [17], who proved the LSI (and T_2) for $\mathbb{P} * \varphi_\sigma$ assuming $\mathbb{E}[e^{aX^2}] < \infty$ holds for some $a > \frac{1}{2\sigma^2}$. They raised a question whether this threshold can be reduced, and our Corollary shows the answer is negative. Indeed, one only needs to noticed that whenever $X \sim \mathbb{P}$ is K-subgaussian it satisfies

$$\mathbb{E}\left[e^{aX^2}\right] < \infty \quad \forall a < \frac{1}{2K^2}\,, \tag{1.5}$$

which is proved in [2, 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. [7] shows when $K < \sigma/2$, $\mathbb{E}\left[W_2^2(P_n * \mathcal{N}(0, \sigma^2 I_d), P * \mathcal{N}(0, \sigma^2 I_d))\right]$ converges with rate $\mathcal{O}\left(\frac{1}{n}\right)$; when $K > \sqrt{2}\sigma$, $\mathbb{E}\left[W_2^2(P_n * \mathcal{N}(0, \sigma^2 I_d), P * \mathcal{N}(0, \sigma^2 I_d))\right]$ 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 between these two. Moreover, [11] shows that $\mathbb{E}\left[W_2(P_n * \mathcal{N}(0, \sigma^2 I_d), P * \mathcal{N}(0, \sigma^2 I_d))\right]$ converges with rate $\mathcal{O}\left(\frac{1}{n^{1/4}}\right)$ for any K and $\sigma > 0$.

Proof idea. In order to prove the convergence rate of smoothed empirical measures, we consider the following quantity: The mutual information

$$I_{\chi^2}(S; Y)$$

where $S \sim \mathbb{P}$, $Y = S + Z$ with $Z \sim \mathcal{N}(0, \sigma^2)$ independent to S. Actually for this $I_{\chi^2}(S; Y)$ we have the following closed-form definition:

$$I_{\chi^2}(S; Y) = \mathbb{E}\left[\chi^2\left(\mathcal{N}(S, \sigma^2 I_d) \middle\| \mathbb{E}\mathcal{N}(\tilde{S}, \sigma^2 I_d)\right)\right],$$

where the first and second expectation are in terms of $S \sim \mathbb{P}$ and $\tilde{S} \sim \mathbb{P}$ respectively, with $S \perp\!\!\!\perp \tilde{S}$.w

According to [7], the convergence rate of smoothed empirical measure under W_2, KL-divergence and the χ^2-divergence is closely related to this $I_{\chi^2}(S;Y)$:

(**Proposition 6** in [7]) If $\mathbb{P}$ is K-subgaussian where $K < \sigma$ 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 [7]) If $I_{\chi^2}(S;Y) = \infty$, then for any $\tau < \sigma$,

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

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 K-subgaussian 1D 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.

Other implications. Results from [7] and the first item of our 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 $O\left(\frac{1}{n}\right)$; and the second item of our Proposition 2 implies that $\mathbb{E}[D_{KL}(P_n * \mathcal{N}(0,\sigma^2 I_d) \| P * \mathcal{N}(0,\sigma^2 I_d))]$ converges with rate $\omega\left(\frac{1}{n}\right)$, and $\mathbb{E}\chi^2(P_n * \mathcal{N}(0,\sigma^2 I_d) \| P * \mathcal{N}(0,\sigma^2 I_d)) = \infty$.

If we suppose the distribution $\mathbb{P}$ is 1D, then we have the following delicate estimation on the convergence of W2 distance:

Theorem 2. *If we know that the distribution $\mathbb{P}$ is 1D in prior, then we have the following two propositions:*

1. *(Lower Bound) For any $K > \sigma > 0$ and $\epsilon > 0$, there exists some K-subgaussian distribution $\mathbb{P}$ such that*

$$\liminf_{n\to\infty} \frac{\mathbb{E}\left[W_2(\mathbb{P}_n * \mathcal{N}(0,\sigma^2 I_d), \mathbb{P} * \mathcal{N}(0,\sigma^2 I_d))\right]}{n^{(\sigma^2+K^2)^2/(4(\sigma^4+K^4))+\epsilon}} > 0.$$

2. *(Upper Bound) Suppose $\mathbb{P}$ is a 1D K-subgaussian random variable ($\mathbb{P}(|x| \geq t | x \sim \mathbb{P}) \leq C \exp\left(-\frac{t^2}{2K^2}\right)$ with $C \geq 1$), and $\mathbb{P}_n$ to be an empirical measure of generated from n samples of $\mathbb{P}$. Then for any $\sigma < K, \epsilon > 0$ we have*

$$\mathbb{E}\left[W_2^2(\mathbb{P} * \mathcal{N}(0,\sigma^2), \mathbb{P}_n * \mathcal{N}(0,\sigma^2))^2\right] = \tilde{\mathcal{O}}\left(n^{-\frac{K^2}{2K^2-\sigma^2}+\epsilon}\right). \tag{1.6}$$

Remark 1. *According to 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

$$\liminf_{n\to\infty} \frac{\mathbb{E}\left[W_2^2(\mathbb{P}_n * \mathcal{N}(0,\sigma^2 I_d), \mathbb{P} * \mathcal{N}(0,\sigma^2 I_d))\right]}{n^{(\sigma^2+K^2)^2/(2(\sigma^4+K^4))+\epsilon}} > 0. \tag{1.7}$$

and upper bound part in Theorem 2 indicates that

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

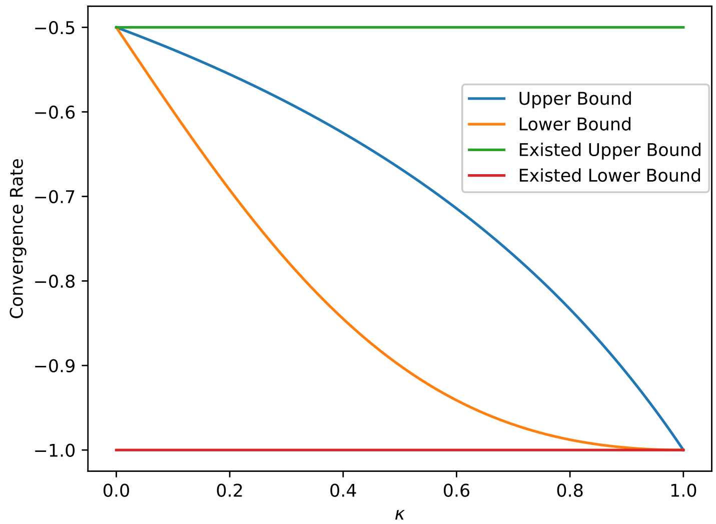
Previously in [7], an upper bound $\mathcal{O}(n^{-1/2})$ and also a lower bound $\omega(n^{-1})$ are demonstrated for cases where $I_{\chi^2}(S;Y) = \infty$. We compare our upper bound and lower bound in (1.6) *and* (1.7)*. The relationship among these bounds are shown in Figure 1. The x-axis is $\kappa = \frac{\sigma^2}{K^2}$ and the y-axis is the convergence rate (the convergence rate is α if we proved a convergence upper or lower bound at $n^{\alpha\pm\epsilon}$ for any $\epsilon > 0$).*

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*

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

Figure 1-1: Relationship of Upper and Lower Bounds of Convergence Rate of $\mathbb{E}\left[W_2^2\right]$ (for $0 < \kappa < 1$)



Remark 2. *From Proposition 1 and 2 and also results from [7], we know that for when $\sigma > K$, the convergence rate is of $\mathcal{O}\left(\frac{1}{n}\right)$. From the above theorem, we know that when $\sigma > K$, the convergence rate is between $\omega\left(\frac{1}{n}\right)$ and $\mathcal{O}\left(\frac{(\log n)^d}{n}\right)$. Hence there is a separation of convergence rate at $\frac{\sigma}{K} < 1$ and $\frac{\sigma}{K} > 1$.*

Remark 3. *Notice that from Theorem 3, we see a dichotomy between the convergence rate of smoothed measures under the W2 distance and under the KL distance. From the lower bound part of Theorem 2 we observe that when $0 < \sigma < K$, the convergence rate under W2 distance must be worse than under KL distance, e.g. $\Omega\left(n^{-(\sigma^2+K^2)^2/(4(\sigma^4+K^4))-\epsilon}\right)$ versus $\tilde{\mathcal{O}}\left(\frac{1}{n}\right)$. This is mainly due to the failure of log-Sobolev inequality for distribution $\mathbb{P} * \mathcal{N}(0, \sigma^2)$ when the subgaussian constant K of $\mathbb{P}$ is greater than σ. (Theorem 1.2 in [17] only applies to cases where $K < \sigma$.)*

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 upper part and the upper bound part of Theorem 2. 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, *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 to be P_X, then $Y \sim P_{Y|X} \circ P_X$.

Furthermore, we use $\mathbf{P}(E)$ to denote the probability of event E, $\mathbb{E}_P[\cdot]$ to denote the expectation with respect to distribution P. We use $A_n = \mathcal{O}(B_n), A_n = \Omega(B_n)$ to denote that $A_n \leq CB_n$ and $A_n \geq CB_n$ for some positive constant C independent of n. We use $A = \tilde{O}(B)$ to denote that $A_n \leq CB_n \cdot \log^l n$ 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.

Chapter 2

Proof of Proposition 1

In this section, we provide proof for Proposition 1. The proof idea is to decompose the integral domain into three subdomains, and we prove that the integral within each subdomain is finite.

Proof. We suppose distribution $\mathbb{P}$ is d-dimensional, and use $\mathcal{N}(0, \sigma^2 I_d)$ to denote the d-dimensional mean-zero multivariate Gaussian distribution with covariance matrix $\sigma^2 I_d$, and $\varphi_{\sigma^2 I_d}(\mathbf{x}) = (\sqrt{2\pi}\sigma)^{-d} \exp\left(-\frac{\|\mathbf{x}\|_2^2}{2\sigma^2}\right)$ to denote its PDF $\mathbf{x} \in \mathbb{R}^d$. 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

$$
\begin{aligned}
I_{\chi^2}(S;Y) &= \mathbb{E}\left[\chi^2\left(\mathcal{N}(S, \sigma^2 I_d) \| \mathbb{E}\mathcal{N}(\tilde{S}, \sigma^2 I_d)\right)\right] \\
&= \mathbb{E}\left[\int_{\mathbb{R}^d} \frac{\varphi_{\sigma^2 I_d}(\mathbf{z} - S)^2}{\mathbb{E}\varphi_{\sigma^2 I_d}(\mathbf{z} - S)} d\mathbf{z} - 1\right] \\
&= (\sqrt{2\pi}\sigma)^{-d}\left[\int_{\mathbb{R}^d} \frac{\exp\left(-\|\mathbf{z} - S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z} - \tilde{S}\|_2^2/(2\sigma^2)\right)} d\mathbf{z}\right] - 1,
\end{aligned}
\tag{2.1}
$$

where $S, \tilde{S} \sim \mathbb{P}$ are $i.i.d.$ Hence we only need to prove that when P is a K-subgaussian distribution with $K < \sigma$,

$$
\mathbb{E}\left[\int_{\mathbb{R}^d} \frac{\exp\left(-\|\mathbf{z} - S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z} - \tilde{S}\|_2^2/(2\sigma^2)\right)} d\mathbf{z}\right] = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{\exp\left(-\|\mathbf{z} - S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z} - \tilde{S}\|_2^2/(2\sigma^2)\right)} dP(S)dz < \infty.
$$

We decompose the integral domain of $(S, \mathbf{z})$ into three sets:

1. $A = \{\|S\|_2 \le 1\}$;

2. $B = \{\|S\|_2 > 1 \text{ and } \|\mathbf{z} - S\|_2 \ge \delta \|S\|_2\}$;

3. $C = \{\|\mathbf{z} - S\|_2 < \delta \|S\|_2\}$.

Since $A \cup B \cup C = \mathbb{R}^d \times \mathbb{R}^d$, and the integrand $\frac{\exp\left(-\|\mathbf{z}-S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z}-\tilde{S}\|_2^2/(2\sigma^2)\right)}$ is always positive, we have

$$\begin{aligned}\mathbb{E}\left[\int_{\mathbb{R}^d} \frac{\exp\left(-\|\mathbf{z}-S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z}-\tilde{S}\|_2^2/(2\sigma^2)\right)} d\mathbf{z}\right] \le\ & \mathbb{E}\left[\int_{\mathbb{R}^d} \mathbf{1}_{(S,z)\in A} \frac{\exp\left(-\|\mathbf{z}-S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z}-\tilde{S}\|_2^2/(2\sigma^2)\right)} d\mathbf{z}\right] \\ &+ \mathbb{E}\left[\int_{\mathbb{R}^d} \mathbf{1}_{(S,z)\in B} \frac{\exp\left(-\|\mathbf{z}-S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z}-\tilde{S}\|_2^2/(2\sigma^2)\right)} d\mathbf{z}\right] \\ &+ \mathbb{E}\left[\int_{\mathbb{R}^d} \mathbf{1}_{(S,z)\in C} \frac{\exp\left(-\|\mathbf{z}-S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z}-\tilde{S}\|_2^2/(2\sigma^2)\right)} d\mathbf{z}\right]\end{aligned}$$

We will proves the finiteness of these three integrals separately.

1. **Part 1:** In this part, we will prove that

$$\begin{aligned}&\mathbb{E}\left[\int_{\mathbb{R}^d} \mathbf{1}_{(S,z)\in A} \frac{\exp\left(-\|\mathbf{z}-S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z}-\tilde{S}\|_2^2/(2\sigma^2)\right)} d\mathbf{z}\right] \\ &= \mathbb{E}\mathbf{1}_{\|S\|_2\le 1}\left[\int_{\mathbb{R}^d} \frac{\exp\left(-\|\mathbf{z}-S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z}-\tilde{S}\|_2^2/(2\sigma^2)\right)} d\mathbf{z}\right] < \infty.\end{aligned} \tag{2.2}$$

We let

$$p_0 = \mathbf{P}(\|S\|_2 \le 1) = \mathbb{E}\left[\mathbf{1}_{\|S\|_2\le 1}\right], \quad S \sim \mathbb{P}.$$

WLOG assume that $p_0 > 0$ (if $p_0 = 0$ then the above formula equals to 0, hence less than infinity). First of all, the denominator in (2.2) has the following lower bound:

$$\begin{aligned}\mathbb{E}\exp\left(-\|\mathbf{z}-\tilde{S}\|_2^2/(2\sigma^2)\right) &\geq \mathbb{E}\left[\mathbf{1}_{\|\tilde{S}\|_2\leq 1}\exp\left(-\|\mathbf{z}-\tilde{S}\|_2^2/(2\sigma^2)\right)\right]\\ &\geq \left(\mathbb{E}\mathbf{1}_{\|\tilde{S}\|\leq 1}\right)\cdot\exp\left(-\frac{(\|\mathbf{z}\|_2+1)^2}{2\sigma^2}\right)\\ &= p_0\exp\left(-\frac{(\|\mathbf{z}\|_2+1)^2}{2\sigma^2}\right)\end{aligned}$$

When $\|S\|_2\leq 1$, we can further upper bound the numerator $\exp\left(-\|\mathbf{z}-S\|_2^2/\sigma^2\right)$ by

$$\exp\left(-\frac{\|\mathbf{z}-S\|_2^2}{\sigma^2}\right)\leq\begin{cases}\exp\left(-(\|\mathbf{z}\|_2-1)^2/\sigma^2\right) & \|\mathbf{z}\|\geq 1,\\ 1 & \|\mathbf{z}\|<1.\end{cases}$$

Therefore,

$$\begin{aligned}&\text{LHS of (2.2)}\\ &\leq\mathbb{E}\left[\mathbf{1}_{\|S\|_2\leq 1}\left(\int_{\|\mathbf{z}\|_2\geq 1}\frac{\exp\left(-(\|\mathbf{z}\|_2-1)^2/\sigma^2\right)}{p_0\exp\left(-\frac{(\|\mathbf{z}\|_2+1)^2}{2\sigma^2}\right)}d\mathbf{z}+\int_{\|\mathbf{z}\|_2<1}\frac{1}{p_0\exp\left(-\frac{(\|\mathbf{z}\|_2+1)^2}{2\sigma^2}\right)}d\mathbf{z}\right)\right]\\ &=\int_{\|\mathbf{z}\|_2\geq 1}\exp\left(-\frac{\|\mathbf{z}\|_2^2}{2\sigma^2}+\frac{3\|\mathbf{z}\|_2}{\sigma^2}-\frac{1}{2\sigma^2}\right)dz+\int_{\|\mathbf{z}\|_2<1}\exp\left(\frac{(\|\mathbf{z}\|_2+1)^2}{2}\right)dz\\ &<\infty,\end{aligned}$$

which proves that inequality (2.2) holds.

2. **Part 2:** In this part, we will prove that for any $\delta>0$,

$$\begin{aligned}&\mathbb{E}\left[\int_{\mathbb{R}^d}\mathbf{1}_{(S,\mathbf{z})\in B}\frac{\exp\left(-\|\mathbf{z}-S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z}-\tilde{S}\|_2^2/(2\sigma^2)\right)}d\mathbf{z}\right]\\ &=\mathbb{E}\left[\mathbf{1}_{\|S\|_2>1}\int_{\|\mathbf{z}-S\|_2\geq\delta\|S\|_2}\frac{\exp\left(-\|\mathbf{z}-S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z}-\tilde{S}\|_2^2/(2\sigma^2)\right)}d\mathbf{z}\right]<\infty.\end{aligned}\tag{2.3}$$

First notice that

$$\{\|S\|_2>1\}\subset\cup_{k=1}^{\infty}A_k,$$

where $A_k=\{S|\|S\|_2\in(k,k+1]\}$. Let l_k to be smallest number of $\mathbb{R}^d$ balls with diameter 2 which can cover the set A_k. Then we have $l_k = \mathcal{O}(k^d)$ (note that here we only need to prove that the integral is finite, hence we ignore the constants). We denote these l_k balls using $A_{k,1}, \cdots, A_{k,l_k}$, where we have $A_k \subset \cup_{i=1}^{l_k} A_{k,i}$. For each $k \geq 1$ and $1 \leq i \leq l_k$, we use $p_{k,i}$ to denote the probability of S in $A_{k,i}$:

$$p_{k,i} = \mathbf{P}(S \in A_{k,i}) = \mathbb{E}\mathbf{1}_{S \in A_{k,i}}, \quad S \sim \mathbb{P}.$$

We notice that for any $S, \tilde{S} \in A_{k,i}$ we have $\|S - \tilde{S}\|_2 \leq 2$, hence

$$\begin{aligned}
&\mathbb{E} \exp\left(-\frac{\|\mathbf{z} - \tilde{S}\|_2^2}{2\sigma^2}\right) \geq \mathbb{E}\left[\mathbf{1}_{\tilde{S} \in A_{k,i}} \exp\left(-\frac{\|\mathbf{z} - \tilde{S}\|_2^2}{2\sigma^2}\right)\right] \\
&\geq p_{k,i} \min_{\tilde{S} \in A_{k,i}} \exp\left(-\frac{\|\mathbf{z} - \tilde{S}\|_2^2}{2\sigma^2}\right) = p_{k,i} \min_{\tilde{S} \in A_{k,i}} \exp\left(-\frac{\|\mathbf{z} - S + (S - \tilde{S})\|_2^2}{2\sigma^2}\right) \\
&\geq p_{k,i} \exp\left(-\frac{(\|\mathbf{z} - S\|_2 + 2)^2}{2\sigma^2}\right).
\end{aligned}$$

We obtain the following upper bound on the expectation in (2.3) specifically for

$S \in A_{k,i}$:

$$
\begin{aligned}
&\mathbb{E}\left[\mathbf{1}_{S\in A_{k,i}} \int_{\|\mathbf{z}-S\|_2 \geq \delta\|S\|_2} \frac{\exp\left(-\|\mathbf{z}-S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z}-\tilde{S}\|_2^2/(2\sigma^2)\right)} d\mathbf{z}\right] \\
&\leq \mathbb{E}\left[\mathbf{1}_{S\in A_{k,i}} \int_{\|\mathbf{z}-S\|_2 \geq \delta\|S\|_2} \frac{\exp\left(-\|\mathbf{z}-S\|_2^2/\sigma^2\right)}{p_{k,i}\exp\left(-\frac{(\|\mathbf{z}-S\|_2+2)^2}{2\sigma^2}\right)} d\mathbf{z}\right] \\
&\leq \mathbb{E}\left[\frac{\mathbf{1}_{S\in A_{k,i}}}{p_{k,i}} \int_{\|\mathbf{z}-S\|_2 \geq \delta\|S\|_2} \exp\left(-\frac{\|\mathbf{z}-S\|_2^2}{\sigma^2} + \frac{(\|\mathbf{z}-S\|_2+2)^2}{2\sigma^2}\right) d\mathbf{z}\right] \\
&= \mathbb{E}\left[\frac{\mathbf{1}_{S\in A_{k,i}}}{p_{k,i}} \int_{\|\mathbf{u}\|_2 \geq \delta\|S\|_2} \exp\left(-\frac{\|\mathbf{u}\|^2}{\sigma^2} + \frac{(\|\mathbf{u}\|_2+2)^2}{2\sigma^2}\right) d\mathbf{u}\right] \\
&\leq \mathbb{E}\left[\frac{\mathbf{1}_{S\in A_{k,i}}}{p_{k,i}} \int_{\|\mathbf{u}\|_2 \geq k\delta} \exp\left(-\frac{\|\mathbf{u}\|^2}{\sigma^2} + \frac{(\|\mathbf{u}\|_2+2)^2}{2\sigma^2}\right) d\mathbf{u}\right] \\
&= \int_{\|\mathbf{u}\|_2 \geq k\delta} \exp\left(-\frac{\|\mathbf{u}\|^2}{\sigma^2} + \frac{(\|\mathbf{u}\|_2+2)^2}{2\sigma^2}\right) d\mathbf{u} \\
&\leq \int_{\|\mathbf{u}\|_2 \geq k\delta} \exp\left(-\frac{\|\mathbf{u}\|_2^2}{\sigma^2} + \frac{\|\mathbf{u}\|_2^2 + \frac{\|\mathbf{u}\|_2^2}{2} + 8 + 4}{2\sigma^2}\right) d\mathbf{u} \\
&= \exp(6/\sigma^2) \int_{\|\mathbf{u}\|_2 \geq k\delta} \exp\left(-\frac{\|\mathbf{u}\|_2^2}{4\sigma^2}\right) d\mathbf{u} \\
&= 2\sigma \exp(6/\sigma^2) \int_{\|\mathbf{u}\|_2 \geq k\delta/(2\sigma)} \exp\left(-\|\mathbf{u}\|_2^2\right) d\mathbf{u},
\end{aligned}
\tag{2.4}
$$

where we use the fact that $\|S\|_2 \geq k$ for $S \in A_k$, and $\mathbb{E}\mathbf{1}_{S\in A_{k,i}} = p_{k,i}$. Moreover, according to changing of variables in integration, there exists some constants C_0 such that

$$
\begin{aligned}
&\int_{\|\mathbf{u}\|_2 \geq k\delta/(2\sigma)} \exp\left(-\|\mathbf{u}\|_2^2\right) d\mathbf{u} \\
&= C_0 \int_{k\delta/(2\sigma)}^{\infty} r^{d-1} \exp\left(-r^2\right) dr \\
&\leq C_0 \int_{k\delta/(2\sigma)}^{\infty} \exp((d-1)r) \exp\left(-r^2\right) dr
\end{aligned}
$$

When $k \leq \frac{2\sigma d}{\delta}$, since the integrand in the above RHS has exponentially decay, there exists a constant C_1 not depending on k, i such that

$$
C_0 \int_{k\delta/(2\sigma)}^{\infty} \exp((d-1)r) \exp\left(-r^2\right) dr \leq C_1
$$

When $r \geq k > \frac{2\sigma d}{\delta}$, we have $(d-1)r - r^2 \leq -r \leq -k$, which indicates that

$$C_0 \int_{k\delta/(2\sigma)}^{\infty} \exp((d-1)r) \exp\left(-r^2\right) dr \leq C_0 \int_{k\delta/(2\sigma)}^{\infty} \exp(-r) dr = C_0 \exp\left(-\frac{k\delta}{2\sigma}\right).$$

Therefore, we obtain that

$$\mathbb{E}\left[\mathbf{1}_{S \in A_{k,i}} \int_{\|\mathbf{z}-S\|_2 \geq \delta\|S\|_2} \frac{\exp\left(-\|\mathbf{z}-S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z}-\tilde{S}\|_2^2/(2\sigma^2)\right)} d\mathbf{z}\right] \leq 2\sigma\exp(6/\sigma^2) C_1$$

for $k \leq \frac{2\sigma d}{\delta}$, and

$$\mathbb{E}\left[\mathbf{1}_{S \in A_{k,i}} \int_{\|\mathbf{z}-S\|_2 \geq \delta\|S\|_2} \frac{\exp\left(-\|\mathbf{z}-S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z}-\tilde{S}\|_2^2/(2\sigma^2)\right)} d\mathbf{z}\right] \leq 2\sigma\exp(6/\sigma^2) C_0 \exp\left(-\frac{k\delta}{2\sigma}\right)$$

for $k > \frac{2\sigma d}{\delta}$. Next noticing that $l_k \leq (2k+3)^d$ and $A_k \subset \cup_{i=1}^{l_k} A_{k,i}$, summing up these expectations for $1 \leq i \leq l_k$, we obtain that for $1 \leq k \leq \frac{2\sigma d}{\delta}$,

$$\mathbb{E}\left[\mathbf{1}_{S \in A_k} \int_{\|\mathbf{z}-S\|_2 \geq \delta\|S\|_2} \frac{\exp\left(-\|\mathbf{z}-S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z}-\tilde{S}\|_2^2/(2\sigma^2)\right)} d\mathbf{z}\right] \leq 2\sigma\exp(6/\sigma^2) C_1 (2k+3)^d,$$

and for $k > \frac{2\sigma d}{\delta}$,

$$\mathbb{E}\left[\mathbf{1}_{S \in A_k} \int_{\|\mathbf{z}-S\|_2 \geq \delta\|S\|_2} \frac{\exp\left(-\|\mathbf{z}-S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z}-\tilde{S}\|_2^2/(2\sigma^2)\right)} d\mathbf{z}\right]$$
$$\leq 2\sigma\exp(6/\sigma^2) C_0 \exp\left(-\frac{k\delta}{2\sigma}\right)(2k+3)^d.$$

Finally summing up for all $k \geq 1$, we obtain that

$$
\begin{aligned}
&\mathbb{E}\left[\mathbf{1}_{\|S\|_2\geq 1}\int_{\|\mathbf{z}-S\|_2\geq\delta\|S\|_2}\frac{\exp\left(-\|\mathbf{z}-S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z}-\tilde{S}\|_2^2/(2\sigma^2)\right)}d\mathbf{z}\right]\\
&=\sum_{k=1}^{\infty}\mathbb{E}\left[\mathbf{1}_{S\in A_k}\int_{\|\mathbf{z}-S\|_2\geq\delta\|S\|_2}\frac{\exp\left(-\|\mathbf{z}-S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z}-\tilde{S}\|_2^2/(2\sigma^2)\right)}d\mathbf{z}\right]\\
&\leq\sum_{1\leq k\leq\frac{2\sigma d}{\delta}}2\sigma\exp(6/\sigma^2)C_1(2k+3)^d+\sum_{k>\frac{2\sigma d}{\delta}}2\sigma\exp(6/\sigma^2)C_0\exp\left(-\frac{k\delta}{2\sigma}\right)(2k+3)^d\\
&<\infty,
\end{aligned}
$$

which proves inequality (2.3).

3. **Part 3:** In this part, we will prove that there exists some $\delta > 0$ (depending on K and $\mathbb{P}$), we have

$$
\mathbb{E}\left[\int_{\mathbb{R}^d}\mathbf{1}_{(S,z)\in C}\frac{\exp\left(-\|\mathbf{z}-S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z}-\tilde{S}\|_2^2/(2\sigma^2)\right)}d\mathbf{z}\right]<\infty, \tag{2.5}
$$

which is equivalent to

$$
\mathbb{E}\left[\int_{\|\mathbf{z}-S\|_2<\delta\|S\|_2}\frac{\exp\left(-\|\mathbf{z}-S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z}-\tilde{S}\|_2^2/(2\sigma^2)\right)}d\mathbf{z}\right]<\infty.
$$

Given the distribution $\mathbb{P}$, first we find constants t, ϵ such that $\mathbf{P}\left(\|\tilde{S}\|_2 \leq t\right) \geq \epsilon$ for $\tilde{S} \sim \mathbb{P}$. This indicates that for any $\delta' > 0$, the following inequality holds for

$\forall z \in \mathbb{R}^d$:

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(-\frac{\|\mathbf{z}-\tilde{S}\|_2^2}{2\sigma^2}\right)\right] \geq \mathbb{E}\left[\mathbf{1}_{\|\tilde{S}\|_2\leq t}\exp\left(-\frac{\|\mathbf{z}-\tilde{S}\|_2^2}{2\sigma^2}\right)\right] \\
&\geq \left(\mathbb{E}\mathbf{1}_{\|\tilde{S}\|_2\leq t}\right)\cdot \min_{\|\tilde{S}\|_2\leq t}\exp\left(-\frac{\|\mathbf{z}-\tilde{S}\|_2^2}{2\sigma^2}\right) \geq \epsilon\cdot\exp\left(-\frac{\left(\|\mathbf{z}\|_2+t\right)^2}{2\sigma^2}\right) \\
&\geq \epsilon\cdot\exp\left(-\frac{\|\mathbf{z}\|_2^2+t^2}{2\sigma^2}-\frac{\delta'\|\mathbf{z}\|_2^2+\frac{t^2}{\delta'}}{2\sigma^2}\right) \\
&= \epsilon\exp\left(-\frac{(1+\delta')t^2}{2\delta'\sigma^2}\right)\cdot\exp\left(-\frac{(1+\delta')\|\mathbf{z}\|_2^2}{2\sigma^2}\right).
\end{aligned}$$

We let

$$\epsilon' = \epsilon\exp\left(-\frac{(1+\delta')t^2}{2\delta'\sigma^2}\right),$$

then given that $\|\mathbf{z}-S\|_2 \leq \delta\|S\|_2$, we have $\|\mathbf{z}\|_2 \leq (1+\delta)\|S\|_2$, and hence

$$\mathbb{E}\left[\exp\left(-\frac{\|\mathbf{z}-\tilde{S}\|_2^2}{2\sigma^2}\right)\right] \geq \epsilon'\exp\left(-\frac{(1+\delta')\|\mathbf{z}\|_2^2}{2\sigma^2}\right) \geq \epsilon'\cdot\exp\left(-\frac{(1+\delta')(1+\delta)^2\|S\|_2^2}{2\sigma^2}\right),$$

This leads to the following estimation on the LHS of (2.5):

$$\begin{aligned}
\text{LHS of (2.5)} &= \mathbb{E}\left[\int_{\|\mathbf{z}-S\|_2<\delta\|S\|_2}\frac{\exp\left(-\|\mathbf{z}-S\|_2^2/\sigma^2\right)}{\mathbb{E}\exp\left(-\|\mathbf{z}-\tilde{S}\|_2^2/(2\sigma^2)\right)}d\mathbf{z}\right] \\
&\leq \frac{1}{\epsilon'}\mathbb{E}\left[\int_{\|\mathbf{z}-S\|_2<\delta\|S\|_2}\exp\left(-\frac{\|\mathbf{z}-S\|_2^2}{\sigma^2}+\frac{(1+\delta')(1+\delta)^2\|S\|_2^2}{2\sigma^2}\right)d\mathbf{z}\right] \\
&\leq \frac{1}{\epsilon'}\mathbb{E}\left[\exp\left(\frac{(1+\delta')(1+\delta)^2}{2\sigma^2}\|S\|_2^2\right)\int_{\mathbb{R}^d}\exp\left(-\frac{\|z-S\|_2^2}{\sigma^2}\right)dz\right] \\
&= \frac{1}{\epsilon'(\sqrt{\pi}\sigma)^d}\mathbb{E}\left[\exp\left(\frac{(1+\delta')(1+\delta)^2}{2\sigma^2}\|S\|_2^2\right)\right].
\end{aligned}$$

Since $K < \sigma$, we have $\frac{K}{\sigma} < 1$. Hence we can choose $\delta, \delta' > 0$ close to 0 such that

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

Then applying Lemma 1 with $\xi = \sigma\sqrt{\frac{1}{(1+\delta)^2(1+\delta')}} > K$ we obtain that LHS of (2.5) $< \infty$. Hence inequality (2.5) is proved.

Finally, combining these three parts of proof, we obtain that the integral in (2.1) is less than infinity, which indicates that $I_{\chi^2}(S;Y) < \infty$. The proof of Proposition 1 is completed. $\square$

In the next, we will present the proof of Lemma 1. This lemma can also be viewed from [2, p. 26].

Lemma 1. *If $S \sim \mathbb{P}$ is K-subgaussian, we have for any $\xi > K$,*

$$\mathbb{E}\left[\exp\left(\frac{\|S\|^2}{2\xi^2}\right)\right] < \infty.$$

Proof. Assume $Z \sim \mathcal{N}(0, \xi^2 I_d)$ independent to S. Then we have

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\frac{\|S\|_2^2}{2\xi^2}\right)\right] \\
&= \frac{1}{\left(\sqrt{2\pi}\xi\right)^d}\mathbb{E}\left[\exp\left(-\frac{\|Z-S\|_2^2}{2\xi^2}\right)\exp\left(\frac{\|S\|_2^2}{2\xi^2}\right)\right] \\
&= \frac{1}{\left(\sqrt{2\pi}\xi\right)^d}\mathbb{E}\left[\exp\left(-\frac{\|Z\|_2^2}{2\xi^2}\right)\exp\left(\frac{S^T Z}{\xi^2}\right)\right] \\
&= \mathbb{E}\left[\exp\left(\frac{S^T Z}{\xi^2}\right)\right] \\
&\le \mathbb{E}\left[\exp\left(\frac{K^2\|Z\|_2^2}{\xi^4}\right)\right] \\
&= \frac{1}{\left(\sqrt{2\pi}\xi\right)^d}\int_{\mathbb{R}^d}\exp\left(-\frac{\xi^2-K^2}{\xi^4}\|\mathbf{z}\|_2^2\right)d\mathbf{z} = \left(\frac{\xi^2}{\xi^2-K^2}\right)^{\frac{d}{2}} < \infty,
\end{aligned} \tag{2.6}$$

which completes the proof of this lemma. $\square$

Chapter 3

Proof of Proposition 2

In this section, we will present a proof of Proposition 2.

Proof. With 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 $I_{\chi^2}(S,Y) = I_{\chi^2}(S',Y')$. Hence we only need to consider S', Z' and Y', which has the property that $Z' \sim \mathcal{N}(0,1)$.)

We construct 1D distribution $\mathbb{P}$ 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

$$p_k = c_1 \exp\left(-\frac{r_k^2}{2K^2}\right), \quad k \geq 1. \tag{3.1}$$

Here we let r_i be a geometrical sequence:

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

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 $p_0 = 1 - \sum_{k=1}^{\infty} p_k > 0$, hence $\mathbb{P}$ is a distribution on $\mathbb{N}$. also P is a K-subgaussian. We can also prove that there exists some $c_1 > 0$ such that for any constant $c > 2$, distribution $\mathbb{P}$ is a K-Subgaussian distribution. The proof of the existence of c_1 is deferred to Section A in appendix.

Remark 4. *If we switch the definition of K-subgaussian of distribution $\mathbb{P}$ from*

$$\forall \alpha : \qquad \mathbb{E}\left[\exp\left(\alpha\left(S - \mathbb{E}[\tilde{S}]\right)\right)\right] \leq \exp\left(\frac{\alpha^2 K^2}{2}\right), \quad S, \tilde{S} \sim \mathbb{P}, S \perp\!\!\!\perp \tilde{S}$$

to

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

then the proof of subgaussian property would be much easier. 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 $k < k_0$, $r_k - \alpha K \leq r_{k_0 - 1} - \alpha K + (k_0 - 1 - k) \leq k_0 - 1 - k$ since $r_{k_0-1} \leq 0$. Hence, we have

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

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

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

We first choose c_1 such that for any c, $\mathbb{P}$ is a K-subgaussian distribution. Then we will specify constant c such that $I_{\chi^2}(S;Y) = \infty$. In the following, we will use $\varphi_\sigma(x)$ to denote the density of 1D Gaussian distribution $\mathcal{N}(0,\sigma^2)$ at x. According to the definition of I_{χ^2}, we have

$$I_{\chi^2}(S;Y) = \int_{\mathbb{R}} \frac{\mathbb{E}\varphi_1^2(z-S)}{\mathbb{E}\varphi_1(z-S)}dz - 1.$$

Hence $I_{\chi^2}(S;Y) = \infty$ is equivalent to

$$\int_{\mathbb{R}} \frac{\mathbb{E}\varphi_{\frac{1}{\sqrt{2}}}(z-S)}{\mathbb{E}\varphi_1(z-S)}dz = \int_{\mathbb{R}} \frac{\sum_{k=0}^{\infty} p_k \varphi_{\frac{1}{\sqrt{2}}}(z-r_k)}{\sum_{k=1}^{\infty} p_k\varphi_1(z-r_k)}dz = \infty.$$

We rewrite the above as

$$\int_{\mathbb{R}} \frac{\sum_{k=0}^{\infty} p_k \varphi_{\frac{1}{\sqrt{2}}}(z-r_k)}{\sum_{k=1}^{\infty} p_k\varphi_1(z-r_k)}dz = \sum_{k=0}^{\infty} \int_{\mathbb{R}} \frac{\varphi_{\frac{1}{\sqrt{2}}}(z-r_k)}{\varphi_1(z-r_k)} \cdot \frac{1}{1+\sum_{j\neq k}\frac{p_j}{p_k}\frac{\varphi_1(z-r_j)}{\varphi_1(z-r_k)}}dz. \tag{3.2}$$

Next, we are going to analyze second term's denominator $1+\sum_{j\neq k}\frac{p_j}{p_k}\frac{\varphi_1(z-r_j)}{\varphi_1(z-r_k)}$ for z close to r_k (when $|z-r_k| \le \delta$ for some $\delta < 1$).

When $j = 0$ and $|z-r_k| \le \delta$, we have

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

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

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

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 $A, B, C \leq 0$ for all k, j. First, notice that because $r_k \geq 1$ or $r_k = 0$ by setting $\delta = \min\left(-\frac{\ell}{2}, \frac{1}{4}\right)$ we have $B, C \leq 0$.

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

Continuing, we obtained that with our choice of c, for $j \neq k, j \geq 1$ and $|z - r_k| \leq \delta$ we have

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

which indicates that $\exists C'$ such that

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

Therefore,

$$\begin{aligned}
&\sum_{k=0}^{\infty} \int_{\mathbb{R}} \frac{\varphi_{\frac{1}{\sqrt{2}}}(z-r_k)}{\varphi_1(z-r_k)} \cdot \frac{1}{1+\sum_{j\neq k} \frac{p_j}{p_k} \frac{\varphi_1(z-r_j)}{\varphi_1(z-r_k)}} dz \\
&\geq \sum_{k=0}^{\infty} \int_{r_k-\delta}^{r_k+\delta} \frac{\varphi_{\frac{1}{\sqrt{2}}}(z-r_k)}{\varphi_1(z-r_k)} \cdot \frac{1}{1+\sum_{j\neq k} \frac{p_j}{p_k} \frac{\varphi_1(z-r_j)}{\varphi_1(z-r_k)}} dz \\
&\geq \left(\int_{-\delta}^{\delta} \frac{\varphi_{\frac{1}{\sqrt{2}}}(z)}{\varphi_1(z)} \right) \cdot \sum_{k=0}^{\infty} \frac{1}{C'} \\
&= \infty
\end{aligned} \tag{3.3}$$

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

Chapter 4

Proof of the Lower Bound Part 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 Simple Bernoulli Distribution Case

We consider Bernoulli distribution $\mathbb{P}_h = (1-p)\delta_0 + p\delta_h$ with $p = \exp\left(-\frac{h^2}{2K^2}\right)$. 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 $\epsilon > 0$,*

$$\sup_h \mathbb{E}\left[W_q(\mathbb{P}_h * \mathcal{N}(0,\sigma^2), \mathbb{P}_{h,n} * \mathcal{N}(0,\sigma^2))\right] = \Omega\left(n^{-\frac{(1+\sigma^2/K^2)^2}{2q(1+\sigma^4/K^4)} - \epsilon}\right),$$

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

Lemma 2. *Suppose two 1D distribution μ, ν with probability density function F_μ, F_ν satisfy $F_\mu(t) \geq F_\nu(t+2)$, then we have*

$$W_p(\mu, \nu) \geq \mathbf{P}(Y \in [t+1, t+2])^{\frac{1}{p}}, \quad Y \sim \nu.$$

Proof. We consider any coupling (X, Y) between μ and ν under the W_p distance, where $X \sim \mu, Y \sim \nu$. Then we have

$$\begin{aligned}
&\mathbf{P}(|X-Y| \geq 1) \geq \mathbf{P}(X \leq t, Y \geq t+1) \geq 1 - \mathbf{P}(X > t) - \mathbf{P}(Y < t+1) \\
&= F_\mu(t) - \mathbf{P}(Y < t+1) \geq F_\nu(t+2) - \mathbf{P}(Y < t+1) = \mathbf{P}(Y \leq t+2) - \mathbf{P}(Y < t+1) \\
&= \mathbf{P}(Y \in [t+1, t+2]).
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
W_p(\mu, \nu)^p &= \inf_{(X,Y) \in \Gamma(\mu,\nu)} \mathbb{E}[\|X-Y\|^p] \\
&\geq \inf_{(X,Y) \in \Gamma(\mu,\nu)} \mathbb{E}[|X-Y|^p \mathbf{1}_{|X-Y| \geq 1}] \\
&\geq \inf_{(X,Y) \in \Gamma(\mu,\nu)} [\mathbf{P}(|X-Y| \geq 1)] \\
&\geq \mathbf{P}(Y \in [t+1, t+2]).
\end{aligned}$$

Therefore, we have $W_p(\mu, \nu) \geq \mathbf{P}(Y \in [t+1, t+2])^{\frac{1}{p}}$. □

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.

We use $\tilde{F}_{n,\sigma}, F_\sigma$ to denote the distribution function of $\mathbb{P}_{h,n} * \mathcal{N}(0, \sigma^2), \mathbb{P}_h * \mathcal{N}(0, \sigma^2)$. Then for $0 < t < h$,

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

where Φ_σ is the distribution function of $\mathcal{N}(0, \sigma^2)$. We let $U_i = \mathbf{1}_{X_k = h}$, then according to Berry-Esseen Theorem, for $V \sim \mathcal{N}(0,1)$, we have

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

When $p_h < 1/2$, we have

$$\begin{aligned}
\mathbb{E}[U_1] &= p_h, \\
\mathrm{Var}[U_1] &= p_h(1-p_h) \geq \frac{1}{2}p_h, \\
\mathbb{E}|U_1 - \mathbb{E}[U_1]|^3 &\leq \mathbb{E}|U_1|^3 = \mathbb{E}[U_1] = p_h.
\end{aligned}$$

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

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

This indicates that

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

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

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

Moreover, 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 $X \sim \mathbb{P}_h * \mathcal{N}(0, \sigma^2)$, we have

$$
\begin{aligned}
&\mathbf{P}(X \in [t, t+2]) \\
&\le 2 \cdot \max_{t' \in [t, t+2]} \left[\frac{1-p_h}{\sqrt{2\pi}\sigma} \exp\left(-\frac{t'^2}{2\sigma^2}\right) + \frac{p_h}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(h-t')^2}{2\sigma^2}\right) \right] \\
&\le \frac{2}{\sqrt{2\pi}\sigma} \cdot \left[\exp\left(-\frac{t^2}{2\sigma^2}\right) + \exp\left(-\frac{h^2}{2K^2} - \frac{(h-t-2)^2}{2\sigma^2}\right) \right], \\
&\mathbf{P}(X \in [t+1, t+2]) \\
&\ge \min_{t' \in [t+1, t+2]} \left[\frac{1-p_h}{\sqrt{2\pi}\sigma} \exp\left(-\frac{t'^2}{2\sigma^2}\right) + \frac{p_h}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(h-t')^2}{2\sigma^2}\right) \right] \\
&\ge \frac{1}{\sqrt{2\pi}\sigma} \cdot \left[\exp\left(-\frac{(t+2)^2}{2\sigma^2}\right) + \exp\left(-\frac{h^2}{2K^2} - \frac{(h-t)^2}{2\sigma^2}\right) \right],
\end{aligned}
\tag{4.1}
$$

where we have use the fact that $t \in (0, h-2)$. Therefore, choosing

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

we notice that $\exists \bar{h} > 0$ such that for $h > \bar{h}$, we have $t \in (0, h-2)$. Moreover, with this choice of t, we have

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

Notice that when h goes to infinity, both t and also $h-t$ goes 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

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

which indicates that

$$
\begin{aligned}
&\mathbf{P}(X \in [t, t+2]) \\
&\leq \frac{4}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(1-\delta)t^2}{2\sigma^2}\right) = \frac{4}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(1-\delta)\left(\frac{1}{2}+\frac{\sigma^2}{2K^2}\right)^2 h^2}{2\sigma^2}\right), \\
&\mathbf{P}(X \in [t+1, t+2]) \\
&\geq \frac{4}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(1+\delta)t^2}{2\sigma^2}\right) = \frac{4}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(1+\delta)\left(\frac{1}{2}+\frac{\sigma^2}{2K^2}\right)^2 h^2}{2\sigma^2}\right)
\end{aligned}
\tag{4.2}
$$

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

$$
n_h = \left\lfloor \frac{1}{18C_1^2} \exp\left(\frac{(1-\delta)\left(\frac{1}{2}+\frac{\sigma^2}{2K^2}\right)^2 h^2}{\sigma^2} - \frac{h^2}{2K^2}\right)\right\rfloor, \tag{4.3}
$$

we have with probability at least $\frac{1}{8} - \frac{1}{\sqrt{2n_h p_h}}$, $\tilde{F}_{n,\sigma}(t) - F_\sigma(t) \geq \mathbf{P}(X \in [t, t+2])$ holds, and this indicates that

$$
\tilde{F}_{n,\sigma}(t) \geq F_\sigma(t+2).
$$

According to Lemma 2, for any $p \geq 1$ we have with probability at least $\frac{1}{8} - \frac{1}{\sqrt{2n_h p_h}}$,

$$
\begin{aligned}
&W_q(\mathbb{P}_h * \mathcal{N}(0,\sigma^2), \mathbb{P}_{h,n_h} * \mathcal{N}(0,\sigma^2))^q \\
&\geq \mathbf{P}(X \in [t+1, t+2]) \\
&\geq C_1 \cdot \exp\left(-\frac{(1+\delta)\left(\frac{1}{2}+\frac{\sigma^2}{2K^2}\right)^2 h^2}{2\sigma^2}\right)
\end{aligned}
$$

Moreover, we notice that the coefficient of h^2 in the exponential term of (4.3)

$$
\frac{(1-\delta)\left(\frac{1}{2}+\frac{\sigma^2}{2K^2}\right)^2}{\sigma^2} - \frac{1}{2K^2} = (1-\delta)\left(\frac{1}{4\sigma^2}+\frac{\sigma^2}{4K^4}\right) - \frac{\delta}{2K^2} \tag{4.4}
$$

is greater than 0 for $0 < \delta < \delta_0$ given the fact that $\sigma < K$. Hence there exists some

h_0 (only depending on K and σ) such that for $h > h_0$, according to (4.3) we have

$$n_h \geq \frac{1}{20C_1^2} \exp\left(\frac{(1-\delta)\left(\frac{1}{2}+\frac{\sigma^2}{2K^2}\right)^2 h^2}{\sigma^2} - \frac{h^2}{2K^2}\right).$$

Also since $\sigma < K$, we have

$$\frac{1}{4\sigma^2} + \frac{\sigma^2}{4K^4} > \frac{1}{2K^2},$$

which indicates that there exists some δ_1 (only depending on K and σ) such that the coefficient in (4.4) satisfies

$$\frac{(1-\delta)\left(\frac{1}{2}+\frac{\sigma^2}{2K^2}\right)^2}{\sigma^2} - \frac{1}{2K^2} > \frac{1}{2K^2}.$$

And hence there exists some h_1 (only depending on K, σ) such that for all $h > h_1$ and $\delta < \frac{\delta_1}{2}$, we have

$$\frac{1}{\sqrt{n_h p_h}} < \frac{1}{16}.$$

Therefore, for all $\delta < \min\{\delta_0, \delta_1/2\}$ and $h > \max\{h_0, h_1\}$, we have with probability $\frac{1}{8} - \frac{1}{16} = \frac{1}{16}$,

$$\begin{aligned}
&W_q(\mathbb{P}_h * \mathcal{N}(0,\sigma^2), \mathbb{P}_{h,n} * \mathcal{N}(0,\sigma^2))^q \\
&\geq C_2 \cdot \exp\left(-\frac{\frac{1}{2}(1+\delta)\left(\frac{1}{2}+\frac{\sigma^2}{2K^2}\right)^2 h^2}{2\sigma^2}\right) \\
&\geq C_2 \cdot \left(20nC_1^2\right)^{-\frac{(1+\delta)\left(\frac{1}{2}+\frac{\sigma^2}{2K^2}\right)^2}{2(1-\delta)\left(\frac{1}{4}+\frac{\sigma^4}{4K^4}\right)-\frac{\delta\sigma^2}{K^2}}}.
\end{aligned}$$

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

$$\mathbb{E}\left[W_q(\mathbb{P}_h * \mathcal{N}(0,\sigma^2), \mathbb{P}_{h,n} * \mathcal{N}(0,\sigma^2))\right] = \Omega\left(n^{-\frac{(1+\sigma^2/K^2)^2}{2q(1+\sigma^4/K^4)}-\epsilon}\right).$$

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

4.2 Proof of the Lower Bound Part of Theorem 2

In this section, we present the proof of the lower bound part of Theorem 2. 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 W2 convergence.

We construct the following discrete distribution

$$\mathbb{P} = \sum_{k=1}^{\infty} p_k \delta_{r_k}, \quad p_k \geq 0, \quad \sum_{k=1}^{\infty} p_k = 1, \tag{4.5}$$

where we choose $r_k = c^{k-1}$ for $k \geq 1$ for some positive constant $c \geq 3$ to be determined later, and

$$p_k = \frac{C}{\sqrt{2\pi}K} \exp\left(-\frac{r_k^2}{2K^2}\right), \quad k \geq 1, \tag{4.6}$$

where C is a constant between 1 and $\sqrt{2K^2\pi}\exp\left(1/2K^2\right)$ such that $\sum_{k=1}^{\infty} p_k = 1$. Then for $X \sim \mathbb{P}$ we have

$$\begin{aligned}\mathbb{E}\left[\exp(\alpha X)\right] &= \sum_{k=1}^{\infty} p_k \exp(\alpha r_k) = \sum_{k=1}^{\infty} \frac{C}{\sqrt{2\pi}K} \exp\left(-\frac{r_k^2}{2K^2} + \alpha r_k\right) \\ &= \sum_{k=1}^{\infty} \frac{C}{\sqrt{2\pi}K} \exp\left(-\frac{1}{2K^2}\left(r_k - \alpha K^2\right)^2\right) \exp\left(\frac{\alpha^2 K^2}{2}\right).\end{aligned} \tag{4.7}$$

Therefore, this distribution is a K-subgaussian.

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

$$t = \frac{1}{2}(c+1)\left(1+\kappa\right) \geq \frac{1}{2}(c+1) \geq 2.$$

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

Proposition 4 (Probability Lower Bound)**.** *There exists some positive constant C_l*

only depending on σ and K such that

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

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

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

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 $c \geq \max\left\{\sqrt{\frac{2}{\kappa}}, \frac{\kappa+3}{1-\kappa}\right\}$, there exists some constant C_u only depending on K and σ such that*

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

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

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

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

1. For $j > k$, we have $r_j^2 \geq r_{k+1}^2 + j - (k+1)$ and also

$$
\begin{aligned}
\frac{(tr_k - r_j + 2)^2}{2\sigma^2} &\geq \frac{(r_{k+1} - tr_k - 2)^2}{2\sigma^2} = \frac{((c-t)r_k - 2)^2}{2\sigma^2} \\
&\geq \frac{(c-t)^2(r_k - 2)^2}{2\sigma^2} = (t^2 - c\kappa - c - c^2\kappa) \cdot \frac{(r_k - 2)^2}{2\sigma^2},
\end{aligned}
$$

after noticing that $c \geq \frac{\kappa+3}{1-\kappa}$ and hence $c - t \geq 1$. Therefore, choosing constant

$$
C_1 = \sum_{j=0}^{\infty} \frac{2\sqrt{2K^2\pi}\exp(1/2K^2)}{\pi\sigma K} \exp\left(-\frac{j}{2K^2}\right) < \infty
$$

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

$$\begin{aligned}
&\sum_{j=k+1}^{\infty} \frac{2C}{\pi\sigma K} \exp\left(-\frac{r_j^2}{2K^2} - \frac{(tr_k - r_j + 2)^2}{2\sigma^2}\right) \\
\leq &\sum_{j=k+1}^{\infty} \frac{2C}{\pi\sigma K} \exp\left(-\frac{j-(k+1)+r_{k+1}^2}{2K^2} - \frac{(t^2 - c\kappa - c - c^2\kappa)(r_k - 2)^2}{2\sigma^2}\right) \\
\leq &\left(\sum_{j=k+1}^{\infty} \frac{2C}{\pi\sigma K} \exp\left(-\frac{j-(k+1)}{2K^2}\right)\right) \cdot \exp\left(-(t^2 - c\kappa - c) \cdot \frac{r_k^2}{2\sigma^2}\right) \\
\leq &\left(\sum_{j=k+1}^{\infty} \frac{2\sqrt{2K^2\pi}\exp(1/2K^2)}{\pi\sigma K} \exp\left(-\frac{j-(k+1)}{2K^2}\right)\right) \cdot \exp\left(-(t^2 - c\kappa - c) \cdot \frac{r_k^2}{2\sigma^2}\right) \\
= &C_1 \exp\left(-(t^2 - c\kappa - c) \cdot \frac{r_k^2}{2\sigma^2}\right).
\end{aligned}$$

2. For $j < k$, noticing that $r_j \leq \frac{r_k}{c}$ and also $c \geq \sqrt{\frac{2}{\kappa}}$, we have

$$(tr_k - r_j)^2 \geq \left(t - \frac{1}{c}\right)^2 r_j^2 \geq (t^2 - \kappa c - c)r_j^2.$$

Therefore, choosing constant $C_2 = \sum_{j=1}^{\infty} \frac{2\sqrt{2K^2\pi}\exp(1/2K^2)}{\pi\sigma K} \exp\left(-\frac{j}{2K^2}\right) < \infty$, we will obtain:

$$\begin{aligned}
&\sum_{j=1}^{k-1} \frac{2C}{\pi\sigma K} \exp\left(-\frac{r_j^2}{2K^2} - \frac{(tr_k - r_j)^2}{2\sigma^2}\right) \\
\leq &\sum_{j=1}^{k-1} \frac{2C}{\pi\sigma K} \exp\left(-\frac{j}{2K^2} - \frac{(t^2 - \kappa c - c)r_k^2}{2\sigma^2}\right) \\
= &\left(\sum_{j=1}^{k-1} \frac{2C}{\pi\sigma K} \exp\left(-\frac{j}{2K^2}\right)\right) \exp\left(-\frac{(t^2 - \kappa c - c)r_k^2}{2\sigma^2}\right) \\
\leq &C_2 \exp\left(-\frac{(t^2 - \kappa c - c)r_k^2}{2\sigma^2}\right) \\
\leq &C_2 \exp\left(-\frac{(t^2 - \kappa c - c)(r_k - 2)^2}{2\sigma^2}\right).
\end{aligned}$$

3. For $j = k$, choosing $C_3 = \frac{2\sqrt{2K^2\pi}\exp\left(1/2K^2\right)}{\pi\sigma K}$, we will obtain:

$$\begin{aligned}\frac{2C}{\pi\sigma K}\exp\left(-\frac{r_k^2}{2K^2}-\frac{(tr_k-r_k)^2}{2\sigma^2}\right) &\le C_l \exp\left(-\left(t^2-\kappa c-c\right)\cdot\frac{r_k^2}{2\sigma^2}\right)\\ &\le C_l \exp\left(-\left(t^2-\kappa c-c\right)\cdot\frac{(r_k-2)^2}{2\sigma^2}\right).\end{aligned}$$

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

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

□

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

Proposition 6. *Suppose $c \ge \frac{\kappa+3}{1-\kappa}$. 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 $\exists k_0 = k_0(\sigma, K, C) > 0$ such that $\forall k \ge k_0$ and n with $np_{k+1} \ge 2048$, with probability at least $\frac{1}{64}$ we have*

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

Proof. First we can write

$$\begin{aligned}F_\sigma(tr_k) &= \sum_{j=1}^{\infty} p_j \Phi_\sigma(tr_k - r_j),\\ \tilde{F}_{n,\sigma}(tr_k) &= \sum_{j=1}^{\infty} \hat{p}_j \Phi_\sigma(tr_k - r_j),\end{aligned}$$

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

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

From assumption $c \geq \frac{\kappa+3}{1-\kappa}$ we know that $c \geq t+1$. Hence for any $j \geq k+1$ we have $|tr_k - r_j| \geq |(c-t)r_k| \geq r_k \geq 1$ and for any $j \leq k$ we have $|tr_k - r_j| \geq (t-1)r_j \geq r_j \geq 1$. According to the upper bound of Gaussian tail function (Proposition 2.1.2 in [15]), when $tr_k - r_j > 0$ we have

$$1 - \Phi_\sigma(tr_k - r_j) \leq \frac{1}{\sqrt{2\pi}} \cdot \frac{\sigma}{|tr_k - r_j|} \exp\left(-\frac{(tr_k - r_j)^2}{2\sigma^2}\right) \leq \sigma \exp\left(-\frac{(tr_k - r_j)^2}{2\sigma^2}\right)$$

and when $tr_k - r_j < 0$ we have

$$\Phi_\sigma(tr_k - r_j) \leq \frac{1}{\sqrt{2\pi}} \cdot \frac{\sigma}{|tr_k - r_j|} \exp\left(-\frac{(tr_k - r_j)^2}{2\sigma^2}\right) \leq \sigma \exp\left(-\frac{(tr_k - r_j)^2}{2\sigma^2}\right).$$

We further notice that

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

Hence adopting Markov inequality we obtained that

$$\mathbf{P}\left(\max_{j\geq 1} |\hat{p}_j - p_j| \leq \frac{4}{\sqrt{n}}\right) \geq \frac{15}{16}. \tag{4.8}$$

In the next, given that $np_{k+1} \geq 2048$, we will provide both a lower bound to $\sum_{j=1}^{k} \hat{p}_j - \sum_{j=1}^{k} p_j$ and also an upper bound to $|\hat{p}_{k+1} - p_{k+1}|$. As for $\sum_{j=1}^{k} \hat{p}_j - \sum_{j=1}^{k} p_j$, we can write it as

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

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

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

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

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

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

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

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

where we use the fact that $np_{k+1} \geq 2048$. As for $|\hat{p}_{k+1} - p_{k+1}|$, according to Bernstein inequality and also noticing that $np_{k+1} \geq 1$, we obtain that

$$\mathbf{P}\left(|\hat{p}_{k+1} - p_{k+1}| \geq 8\sqrt{\frac{p_{k+1}}{n}}\right) \leq 2\exp\left(-\frac{64np_{k+1}}{\frac{2}{3}\cdot 8\sqrt{np_{k+1}} + 2np_{k+1}(1-p_{k+1})}\right)$$
$$\leq 2\exp\left(-\frac{64}{16/3+2}\right) \leq 2\exp(-8) \leq \frac{1}{64}.$$

Therefore, if $n \geq 2048/p_{k+1}$, according to (4.8), with probability at least $\frac{1}{64}$ we have

$$\begin{aligned} &\tilde{F}_{n,\sigma}(2r_k) - F_\sigma(2r_k) \\ &\geq \sqrt{\frac{p_{k+1}}{2n}} - \frac{4\sigma}{\sqrt{n}}\sum_{j=1}^{k}\exp\left(-\frac{(tr_k - r_j)^2}{2\sigma^2}\right) - \frac{4\sigma}{\sqrt{n}}\sum_{j=k+2}^{\infty}\exp\left(-\frac{(tr_k - r_j)^2}{2\sigma^2}\right) \\ &\quad - 8\sigma\sqrt{\frac{p_{k+1}}{n}}\exp\left(-\frac{(tr_k - r_{k+1})^2}{2\sigma^2}\right). \end{aligned} \tag{4.9}$$

Additionally, we have

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

and since for any $j \geq k+2$, $r_j - tr_k \geq j-(k+2) + r_{k+2} - tr_k \geq j-(k+2)+(t-1)t_k$, we have

$$\begin{aligned} &\sum_{j=k+2}^{\infty}\exp\left(-\frac{(tr_k - r_j)^2}{2\sigma^2}\right) \\ &\leq \left(\sum_{j=k+2}^{\infty}\exp\left(-\frac{j-(k+2)}{2\sigma^2}\right)\right)\cdot\exp\left(-\frac{(t-1)^2 r_k^2}{2\sigma^2}\right) \\ &\leq C_4\exp\left(-\frac{(t-1)^2 r_k^2}{2\sigma^2}\right), \end{aligned}$$

where C_4 is a constant only depending on σ, K and c. We also notice that $\frac{(tr_k - r_{k+1})^2}{2\sigma^2} \geq \frac{r_k^2}{2\sigma^2}$ using the fact that $c \geq t+1$, and that

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

using the fact that

$$2c^2\kappa + c^2\kappa^2 \le c^2\kappa^2 + c^2 + \kappa^2 + 1 - 2c - 2\kappa + 2c^2\kappa = (2t-2)^2.$$

Hence we have

$$\begin{aligned}
&\frac{4\sigma}{\sqrt{n}}\sum_{j=1}^{k}\exp\left(-\frac{(tr_k - r_j)^2}{2\sigma^2}\right) + \frac{4\sigma}{\sqrt{n}}\sum_{j=k+2}^{\infty}\exp\left(-\frac{(tr_k - r_j)^2}{2\sigma^2}\right)\\
&\quad + 8\sigma\sqrt{\frac{p_{k+1}}{n}}\exp\left(-\frac{(tr_k - r_{k+1})^2}{2\sigma^2}\right)\\
&\le 4\sqrt{\frac{p_{k+1}}{n}}\cdot\sigma\left((C_4+k)\exp\left(-\frac{c^2\kappa^2 r_k^2}{8\sigma^2}\right) + \exp\left(-\frac{r_k^2}{2\sigma^2}\right)\right).
\end{aligned}$$

Since $r_k = c^{k-1}$ with $c \ge 2$, there exists some constant k_0 only depending on K, σ and c such that for any $k \ge k_0$, we have

$$\sigma\left((C_4+k)\exp\left(-\frac{c^2\kappa^2 r_k^2}{8\sigma^2}\right) + \exp\left(-\frac{r_k^2}{2\sigma^2}\right)\right) \le \frac{1}{4\sqrt{2}} - \frac{1}{8}$$

Bringing this result to (4.9), we will obtain that for any $k \ge k_0$,

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

holds. This completes the proof of this proposition. $\square$

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

Proof of the Lower Bound Part of Theorem 2. We choose

$$n = \left\lfloor \frac{1}{4C_u^2}\exp\left((t^2 - c\kappa - c)\cdot\frac{(r_k-2)^2}{\sigma^2} - \frac{c^2 r_k^2}{2K^2}\right)\right\rfloor. \tag{4.10}$$

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

$$np_{k+1} \ge 2048.$$

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

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

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

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

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

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

Therefore we have

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

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

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

Since we have

$$\lim_{k \to \infty} \frac{r_k + 2}{r_k - 2} = \lim_{k \to \infty} \frac{r_k + 2}{r_k} = 1$$

we would obtain that for any $\epsilon', c > \max\left\{\sqrt{\frac{2}{\kappa}}, \frac{\kappa+3}{1-\kappa}\right\}$,

$$\liminf_{n \to \infty} \frac{\mathbb{E}[W_2(\mathbb{P} * \mathcal{N}(0, \sigma^2), \mathbb{P}_n * \mathcal{N}(0, \sigma^2))]}{n^{\frac{(t^2 - c\kappa - c)/(4\sigma^2)}{(t^2 - c\kappa - c)/\sigma^2 - c^2/(2K^2)} + \epsilon}} > 0.$$

Choosing c larger enough, and remembering that $t = \frac{1}{2}(1+\kappa)(1+c)$, we would obtain that for any $\epsilon > 0$,

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{\mathbb{E}[W_2(\mathbb{P} * \mathcal{N}(0,\sigma^2), \mathbb{P}_n * \mathcal{N}(0,\sigma^2))]}{n^{\frac{(1+\kappa)^2}{4(1+\kappa^2)}+\epsilon}} \\
&= \liminf_{n\to\infty} \frac{\mathbb{E}[W_2(\mathbb{P} * \mathcal{N}(0,\sigma^2), \mathbb{P}_n * \mathcal{N}(0,\sigma^2))]}{n^{\frac{(1+\kappa)^2 c^2/16}{(1+\kappa)^2 c^2/4 - 4\cdot\kappa^2 c^2}+\epsilon}} \\
&> 0.
\end{aligned}$$

Replacing κ with $\frac{\sigma^2}{K^2}$, the proof of the lower bound part of Theorem 2 is completed. $\square$

Chapter 5

Proof of the Upper Bound Part of Theorem 2

We divide the proof into four parts.

In the first part, we provide some useful propositions regarding the distribution after convolving with Gaussian:

1. The PDF of the distribution after convolving with Gaussian can be uniformly upper bounded and strictly lower bounded by 0 (Proposition 7);

2. The PDF of the distribution after convolving with Gaussian does not deviate too much in the neighborhood (Proposition 8);

3. The distribution after convolving with Gaussian is still a sub-Gaussian distribution if the original distribution is a sub-Gaussian distribution (Proposition 9).

In the second part, we provide some propositions regarding the bounds on PDFs, CDFs and Wasserstein distance of distributions:

1. For 1D sub-Gaussian distribution $\mathbb{P}$, the CDF of $\mathbb{P}*\mathcal{N}(0,\sigma^2)$ can be lower/upper bounded using the PDF of $\mathbb{P}*\mathcal{N}(0,\sigma^2)$ (Proposition 10).

2. Considering two 1D-distributions $\mathbb{P},\mathbb{Q}$ with PDFs ρ_p,ρ_q and CDFs F_p,F_q. Then

the 2-Wasserstein distribution between $\mathbb{P}$ and $\mathbb{Q}$ can be written as an integral of $\rho_p(t)\left|F_q^{-1}(F_p(t)-t)\right|^2$ (Proposition 11).

3. Considering two 1D-distributions $\mathbb{P}, \mathbb{Q}$. If the ratio between the supreme of the difference between CDFs of $\mathbb{P}, \mathbb{Q}$ and the infimum of PDF of $\mathbb{P}$ in the neighborhood of t can be upper bounded, then we can obtain an upper bound on $\left|F_q^{-1}(F_p(t)-t)\right|$ (Proposition 12).

4. Considering two 1D-distributions $\mathbb{P}, \mathbb{Q}$. Suppose CDFs of $\mathbb{P} * \mathcal{N}(0, \sigma^2), \mathbb{Q} * \mathcal{N}(0, \sigma^2)$ to be $F_{p,\sigma}, F_{q,\sigma}$. Then if distribution $\mathbb{P}, \mathbb{Q}$ are both sub-Gaussian distributions, then for any $R > 0$, $\left|F_{q,\sigma}^{-1}(F_{p,\sigma}(t)) - t\right|$ can be uniformly upper bounded for those $t \in [-R, R]$ (Proposition 13).

In the third part, we provide some useful propositions with respect to the empirical measures after smoothing:

1. For 1D-distribution $\mathbb{P}$ and its empirical version $\mathbb{P}_n$, we use $F_\sigma, \tilde{F}_{n,\sigma}$ to denote the CDFs of $\mathbb{P} * \mathcal{N}(0, \sigma^2), \mathbb{P}_n * \mathcal{N}(0, \sigma^2)$. We provide a strong uniform upper bound for $\left|F_\sigma(\cdot) - \tilde{F}_{n,\sigma}(\cdot)\right|$ (Proposition 14). If without smoothing, we can also provide such a strong uniform upper bound for difference between CDFs of original and empirical measures (Lemma 5).

2. For 1D-distribution $\mathbb{P}$ and its empirical version $\mathbb{P}_n$, we use $F_\sigma, \tilde{F}_{n,\sigma}$ to denote the CDFs of $\mathbb{P} * \mathcal{N}(0, \sigma^2), \mathbb{P}_n * \mathcal{N}(0, \sigma^2)$. Then if distribution $\mathbb{P}$ is sub-Gaussian, with high probability for any $R > 0$, $\left|F_{q,\sigma}^{-1}(F_{p,\sigma}(t)) - t\right|$ can be uniformly upper bounded for those $t \in [-R, R]$ (Proposition 15).

Finally in the last part, to upper bound the squared 2-Wasserstein distance between $\mathbb{P} * \mathcal{N}(0, \sigma^2)$ and $\mathbb{P}_n * \mathcal{N}(0, \sigma^2)$, we write it as an integral $\rho_\sigma(t)\left|\tilde{F}_{n,\sigma}^{-1}(F_\sigma(t) - t)\right|^2$, where ρ_σ is the PDF of $\mathbb{P} * \mathcal{N}(0, \sigma^2)$. We define $a(t) \triangleq \sqrt{2\sigma^2 \log \frac{1}{\sqrt{2\pi}\sigma\rho_\sigma(t)}} \in [0, \infty)$ and we divide the integral domain of t into three parts based on the value of $a(t)$. And we will bound the integral within each part individually.

Proposition 7. *Suppose ρ_σ is the PDF of $\mathbb{P} * \mathcal{N}(0, \sigma^2)$ for some 1D distribution $\mathbb{P}$, then for $\forall t \in \mathbb{R}$ we have*

$$0 < \rho_\sigma(t) \le \frac{1}{\sqrt{2\pi}\sigma}.$$

Proof. Suppose $\eta_\sigma(\cdot) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\cdot)^2}{2\sigma^2}\right)$ to be the PDF of $\mathcal{N}(0, \sigma^2)$, and let $X \sim \mathbb{P}$. Then for any $t \in \mathbb{R}$ we have

$$\rho_\sigma(t) = \mathbb{E}\left[\eta_\sigma(t - X)\right] = \mathbb{E}\left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(X-t)^2}{2\sigma^2}\right)\right] \le \mathbb{E}\left[\frac{1}{\sqrt{2\pi}\sigma}\right] = \frac{1}{\sqrt{2\pi}\sigma}.$$

Moreover, since $\lim_{K \to \infty} \mathbf{P}(|X| \le K) = \mathbf{P}(X \in \mathbb{R}) = 1$, there exists some K such that $\mathbf{P}(|X| \le K) > 0$. Hence,

$$\begin{aligned}\rho_\sigma(t) &= \mathbb{E}\left[\eta_\sigma(t - X)\right] \ge \mathbb{E}\left[\mathbf{1}_{|X| \le K} \eta_\sigma(t - X)\right] \\ &\ge \mathbf{P}(|X| \le K) \cdot \exp\left(-\frac{(|X| + t)^2}{2\sigma^2}\right) > 0.\end{aligned}$$

□

Proposition 8. *Suppose ρ_σ to be the density function of $\mathbb{P} * \mathcal{N}(0, \sigma^2)$. If for some t and $a \ge 0$ we have $\rho_\sigma(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{a^2}{2\sigma^2}\right)$, then for any δ we have*

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(a + |\delta| + 4\sigma)^2}{2\sigma^2}\right) \le \rho_\sigma(t + \delta) \le \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\max\{0, a - |\delta| - 4\sigma\}^2}{2\sigma^2}\right).$$

Proof. We first prove the upper bound. WLOG, we assume $t = 0$ and $\delta \ge 0$. Then according to the assumption we have

$$\rho_\sigma(0) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{a^2}{2\sigma^2}\right).$$

We use $\eta_\sigma(\cdot) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\cdot)^2}{2\sigma^2}\right)$ to denote the PDF of $\mathcal{N}(0, \sigma^2)$, and let $X \sim \mathbb{P}$. Then noticing that $\eta_\sigma(\cdot)$ is symmetric with respect to 0, we have for $\forall 0 \le r \le a$,

$$\begin{aligned}\rho_\sigma(0) &= \mathbb{E}\left[\eta_\sigma(-X)\right] = \mathbb{E}\left[\eta_\sigma(X)\right] \ge \mathbb{E}\left[\eta_\sigma(X) \mathbf{1}_{|X| \le r}\right] \\ &\ge \mathbf{P}(|X| \le r) \cdot \min_{|x| \le r} \eta_\sigma(x) = \mathbf{P}(|X| \le r) \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{r^2}{2\sigma^2}\right).\end{aligned}$$

Therefore, we obtain that

$$\mathbf{P}(|X| \le r) \le \exp\left(-\frac{a^2 - r^2}{2\sigma^2}\right).$$

Substituting r with $r+\delta$ and noticing that $\left\{x \big| |x-\delta| \le r\right\} \subset \left\{x \big| |x| \le r+\delta\right\}$, we have

$$\mathbf{P}(|X-\delta| \le r) \le \mathbf{P}(|X| \le r+\delta) \le \exp\left(-\frac{a^2 - (r+\delta)^2}{2\sigma^2}\right).$$

Next, we notice that

$$\begin{aligned}
\rho_\sigma(\delta) &= \mathbb{E}[\eta_\sigma(\delta - X)] = \mathbb{E}[\eta_\sigma(X-\delta)] \\
&= \mathbb{E}\left[\eta_\sigma(X-\delta)\mathbf{1}_{|X-\delta| \le a-\delta}\right] + \mathbb{E}\left[\eta_\sigma(X-\delta)\mathbf{1}_{|X-\delta| > a-\delta}\right] \\
&= \int_0^{a-\delta} \eta_\sigma(r) d\mathbf{P}(|X-\delta| \le r) + \mathbb{E}\left[\eta_\sigma(X-\delta)\mathbf{1}_{|X-\delta| > a-\delta}\right] \\
&\le \int_0^{a-\delta} \eta_\sigma(r) d\mathbf{P}(|X-\delta| \le r) + \sup_{|x-\delta|>1-\delta} \eta_\sigma(x-\delta) \\
&= \int_0^{a-\delta} \eta_\sigma(r) d\mathbf{P}(|X-\delta| \le r) + \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(a-\delta)^2}{2\sigma^2}\right),
\end{aligned}$$

where $d\mathbf{P}(|X-\delta| \le r)$ denotes the differential form with respect to r. Adopting integration by part to the first item, we obtain that

$$\begin{aligned}
&\int_0^{a-\delta} \eta_\sigma(r) d\mathbf{P}(|X-\delta| \le r) \\
&\le \eta_\sigma(a-\delta)\mathbf{P}(|X-\delta| \le a-\delta) - \eta_\sigma(0)\mathbf{P}(|X-\delta| \le 0) - \int_0^{a-\delta} \mathbf{P}(|X-\delta| \le r) d\eta_\sigma(r) \\
&\le \eta_\sigma(a-\delta) + \int_0^{a-\delta} \mathbf{P}(|X-\delta| \le r) \cdot \frac{1}{\sqrt{2\pi}\sigma} \cdot \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) dr \\
&= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(a-\delta)^2}{2\sigma^2}\right) + \int_0^{a-\delta} \mathbf{P}(|X-\delta| \le r) \cdot \frac{1}{\sqrt{2\pi}\sigma} \cdot \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) dr.
\end{aligned}$$

Hence we have

$$
\begin{aligned}
&\rho_\sigma(\delta) \\
&\leq \int_0^{a-\delta} \mathbf{P}(|X-\delta| \leq r) \cdot \frac{1}{\sqrt{2\pi}\sigma} \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) dr + \frac{2}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(a-\delta)^2}{2\sigma^2}\right) \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_0^{a-\delta} \frac{r}{\sigma^2} \exp\left(-\frac{a^2-(r+\delta)^2+r^2}{2\sigma^2}\right) dr + \frac{2}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(a-\delta)^2}{2\sigma^2}\right) \\
&\leq \frac{1}{\sqrt{2\pi}\sigma} \int_0^{a-\delta} \frac{r}{\sigma^2} \exp\left(-\frac{(a-\delta)^2}{2\sigma^2}\right) dr + \frac{2}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(a-\delta)^2}{2\sigma^2}\right),
\end{aligned}
$$

where in the last inequality we use the fact that

$$
a^2-(r+\delta)^2+r^2 = a^2-2r\delta+\delta^2 \geq a^2-2a\delta+\delta^2 = (a-\delta)^2, \quad \forall 0 \leq r \leq a-\delta.
$$

Further we notice

$$
\begin{aligned}
&\frac{1}{\sqrt{2\pi}\sigma} \int_0^{a-\delta} \frac{r}{\sigma^2} \exp\left(-\frac{(a-\delta)^2}{2\sigma^2}\right) dr + \frac{2}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(a-\delta)^2}{2\sigma^2}\right) \\
&= \frac{1}{\sqrt{2\pi}\sigma} \left(\frac{(a-\delta)^2}{2\sigma^2} + 2\right) \exp\left(-\frac{(a-\delta)^2}{2\sigma^2}\right) \\
&= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(a-\delta)^2}{2\sigma^2} + \log\left(1+\frac{(a-\delta)^2}{4\sigma^2}\right) + \log 2\right).
\end{aligned}
$$

If we let $\xi = \frac{a-\delta}{\sigma}$ and assume $\xi \geq 4$, we have

$$
\begin{aligned}
&-\frac{(a-\delta)^2}{2\sigma^2} + \log\left(1+\frac{(a-\delta)^2}{4\sigma^2}\right) + \log 2 = -\frac{\xi^2}{2} + \log\left(1+\frac{\xi^2}{4}\right) + \log 2 \\
&\leq -\frac{\xi^2}{2} + \log\left(1+\frac{\xi}{2}\right)^2 + 1 \leq -\frac{\xi^2}{2} + 2\cdot\frac{\xi}{2} + 1 = -\frac{\xi^2}{2} + \xi + 1 \leq -\frac{(\xi-4)^2}{2}.
\end{aligned}
$$

Hence we obtain that

$$
\begin{aligned}
\rho_\sigma(\delta) &\leq \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(a-\delta)^2}{2\sigma^2} + \log\left(1+\frac{(a-\delta)^2}{4\sigma^2}\right) + \log 2\right) \\
&\leq \exp\left(-\frac{(\xi-4)^2}{2}\right) = \exp\left(-\frac{(a-\delta-4\sigma)^2}{2\sigma^2}\right)
\end{aligned}
$$

Moreover, when $\alpha-\delta \leq 4\sigma$, Proposition 7 indicates that $\rho_\sigma(\delta) \leq \frac{1}{\sqrt{2\pi}\sigma}$. Therefore, we obtain the upper bound

$$\rho_\sigma(\delta) \le \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\max\{0, a-\delta-4\sigma\}^2}{2\sigma^2}\right).$$

Next, we consider the lower bound. Proposition 7 indicates that $\rho_\sigma(t+\delta) \le \frac{1}{\sqrt{2\pi}\sigma}$, hence we can let $\rho_\sigma(t+\delta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{b^2}{2\sigma^2}\right)$ with $b \ge 0$. Then the upper bound we just proved indicates that

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{a^2}{2\sigma^2}\right) = \rho_\sigma(t) \le \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\max\{0, b-\delta-4\sigma\}^2}{2\sigma^2}\right),$$

which indicates that $\max\{0, b-\delta-4\sigma\} \le a$. Hence we have $b \le a+\delta+4\sigma$, and

$$\rho_\sigma(t+\delta) \ge \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(a+|\delta|+4\sigma)^2}{2\sigma^2}\right).$$

□

Proposition 9. *Suppose $\sigma < K$, and for $X \sim \mathbb{P}$ we have*

$$\mathbf{P}(|X| \ge t) \le C \exp\left(-\frac{t^2}{2K^2}\right),$$

then for $Y \sim \mathbb{P} * \mathcal{N}(0, \sigma^2)$, we have

$$\mathbf{P}(|Y| \ge t) \le \left(C + \frac{1}{\sqrt{2\pi}\sigma}\right) \exp\left(-\frac{t^2}{8K^2}\right)$$

Proof. First we notice that for $Y \sim \mathbb{P} * \mathcal{N}(0, \sigma^2)$, we can write it as

$$Y = X + Z, \qquad X \sim \mathbb{P}, \quad Z \sim \mathcal{N}(0, \sigma^2), \quad X \perp\!\!\!\perp Z$$

Since $\sigma < K$ and $\mathbf{P}(|X| \geq t) \leq C \exp\left(-\frac{t^2}{2K^2}\right)$, we obtain

$$
\begin{aligned}
\mathbf{P}(|Y| \geq t) &= \mathbf{P}(|X+Z| \geq t) \\
&\leq \mathbf{P}(|X| \leq t/2) + \mathbf{P}(|Z| \leq t - t/2) \\
&\leq C \exp\left(-\frac{t^2}{8K^2}\right) + \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{t^2}{8\sigma^2}\right) \\
&\leq \left(C + \frac{1}{\sqrt{2\pi}\sigma}\right) \exp\left(-\frac{t^2}{8K^2}\right).
\end{aligned}
$$

□

Proposition 10. *We denote the CDF, PDF of $\mathbb{P} * \mathcal{N}(0, \sigma^2)$ as F_σ, ρ_σ respectively. Suppose there exist constants $C, K > 0$ such that for $\forall r \geq 0$,*

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

Then for any $\varepsilon > 0$, $\exists M = M(\varepsilon, \sigma, K, C)$ such that

$$
\begin{aligned}
1 - F_\sigma(r) &\leq M \rho_\sigma(r)^{\frac{\sigma^2}{K^2+\varepsilon}}, \quad \forall r \geq 0, \\
F_\sigma(r) &\leq M \rho_\sigma(r)^{\frac{\sigma^2}{K^2+\varepsilon}}, \quad \forall r < 0.
\end{aligned}
$$

First we present the following lemma:

Lemma 3. *Suppose Φ_σ to be the CDF of Gaussian distribution $\mathcal{N}(0, \sigma^2)$, then we have*

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

Proof. Since we have $\Phi_\sigma(l) = 1 - \Phi_\sigma(l)$ 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 (Proposition 2.1.2 in [15]), we have for $l \geq \sigma$,

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

For $0 \le l \le \sigma$, we have

$$1 - \Phi_\sigma(l) \le 1 - \frac{1}{2} = \frac{1}{2}, \quad \exp\left(-\frac{l^2}{2\sigma^2}\right) \ge \exp(-1/2) \ge \frac{1}{2},$$

which indicates that

$$1 - \Phi_\sigma(l) \le \frac{1}{2} \le \exp\left(-\frac{l^2}{2\sigma^2}\right).$$

Hence for $\forall l \ge 0$,

$$1 - \Phi_\sigma(l) \le \exp\left(-\frac{l^2}{2\sigma^2}\right).$$

$\square$

Proof of Proposition 10. We only prove this results for $r \ge 0$, as the proof of $r \le 0$ is similar. In the following, we use ρ to denote the PDF of $\mathbb{P}$ (which can be a generalized function on $\mathbb{R}$), and Φ_σ to denote the CDF of $\mathcal{N}(0, \sigma^2)$. Then we can write

$$\begin{aligned} 1 - F_\sigma(r) &= \int_{-\infty}^{\infty} \rho(t)(1 - \Phi_\sigma(r-t))dt, \\ \rho_\sigma(r) &= \int_{-\infty}^{\infty} \rho(t) \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(r-t)^2}{2\sigma^2}\right) dt. \end{aligned} \tag{5.1}$$

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

$$R_0 = K\sqrt{2(\log C + 1)},$$

we will obtain that

$$\mathbf{P}(|X| \ge R_0) \le C \exp(-\log C - 1) = \frac{1}{e} < \frac{1}{2}$$

and hence $\mathbf{P}(|X| \le R_0) \ge \frac{1}{2}$. In the following, we will discuss cases where $0 \le r \le R_0$ and $r > R_0$ separately.

If $0 \le r \le R_0$, then we have

$$\rho_\sigma(r) \ge \int_{-R_0}^{R_0} \rho(t) \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(r-t)^2}{2\sigma^2}\right) dt$$
$$\ge \frac{1}{\sqrt{2\pi}\sigma} \mathbf{P}(|X| \le R_0) \cdot \min_{0 \le r \le R_0, t \in [-R_0, R_0]} \exp\left(-\frac{(r-t)^2}{2\sigma^2}\right) = \frac{1}{2\sqrt{2\pi}\sigma} \exp\left(-\frac{2R_0^2}{\sigma^2}\right).$$

We further notice that $1 - F_\sigma(r) \le 1$. Hence for any $\varepsilon > 0$, if choosing $M_1 = \left(\frac{1}{2\sqrt{2\pi}\sigma} \exp\left(-\frac{2R_0^2}{\sigma^2}\right)\right)^{-\frac{\sigma^2}{K^2+\varepsilon}}$, we will have

$$1 - F_\sigma(r) \le 1 \le M_1 \rho_\sigma(r)^{\frac{\sigma^2}{K^2+\varepsilon}}, \quad \forall r \in [0, R_0].$$

Next, we consider cases where $r > R_0$. According to the assumption, we have

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

which indicates that

$$\begin{aligned}
1 - F_\sigma(r) &= \int_{-\infty}^{r} \rho(t)(1 - \Phi_\sigma(r-t))dt + \int_{r}^{\infty} \rho(t)(1 - \Phi_\sigma(r-t))dt \\
&\le \int_{-\infty}^{r} \rho(t)(1 - \Phi_\sigma(r-t))dt + \int_{r}^{\infty} \rho(t)dt \\
&\le \int_{-\infty}^{r} \rho(t)(1 - \Phi_\sigma(r-t))dt + \mathbb{P}(X \ge r) \\
&\le \int_{-\infty}^{r} \rho(t)(1 - \Phi_\sigma(r-t))dt + C \exp\left(-\frac{r^2}{2K^2}\right).
\end{aligned}$$

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

$$\begin{aligned}
&1 - F_\sigma(r) \\
&\le \int_{-\infty}^{r} \rho(t) \exp\left(-\frac{(r-t)^2}{2\sigma^2}\right) dt + C \exp\left(-\frac{r^2}{2K^2}\right) \\
&\le \int_{-\infty}^{\infty} \rho(t) \exp\left(-\frac{(r-t)^2}{2\sigma^2}\right) dt + C \exp\left(-\frac{r^2}{2K^2}\right) \\
&= \sqrt{2\pi}\sigma \cdot \rho_\sigma(r) + C \exp\left(-\frac{r^2}{2K^2}\right).
\end{aligned}$$

Moreover, since $\mathbf{P}(|X| \leq R_0) \geq \frac{1}{2}$ we also have

$$\rho_\sigma(r) \geq \int_{-R_0}^{R_0} \rho(t) \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(r-t)^2}{2\sigma^2}\right) dt \geq \frac{1}{2\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(r+R_0)^2}{2\sigma^2}\right).$$

Given any $\varepsilon > 0$, according to AM-GM inequality, we have

$$\frac{(r+R_0)^2}{2\sigma^2} = \frac{r^2}{2\sigma^2} + \frac{R_0^2}{2\sigma^2} + \frac{2rR_0}{2\sigma^2} \leq \frac{r^2}{2\sigma^2} \cdot \frac{K^2+\epsilon}{K^2} + \frac{R_0^2}{2\sigma^2} \cdot \frac{\epsilon + K^2}{\epsilon},$$

which indicates that

$$\exp\left(-\frac{(r+R_0)^2}{2\sigma^2}\right) \geq \exp\left(-\frac{R_0^2}{2\sigma^2} \cdot \frac{\epsilon+K^2}{\epsilon}\right) \cdot \exp\left(-\frac{r^2}{2\sigma^2} \cdot \frac{K^2+\epsilon}{K^2}\right).$$

Therefore, choosing

$$M_2 = M_2(\varepsilon, \sigma, K, C) = C\left(2\sqrt{2\pi}\sigma\right)^{\frac{\sigma^2}{K^2+\varepsilon}} \exp\left(\frac{R_0^2}{2\epsilon}\right),$$

we will have

$$\begin{aligned}
&M_2 \cdot \rho_\sigma(r)^{\frac{\sigma^2}{K^2+\varepsilon}} \\
&\geq M_2 \cdot \left(\frac{1}{2\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(r+R_0)^2}{2\sigma^2}\right)\right)^{\frac{\sigma^2}{K^2+\varepsilon}} \\
&\geq C \exp\left(\frac{R_0^2}{2\epsilon}\right) \cdot \left(\exp\left(-\frac{R_0^2}{2\sigma^2} \cdot \frac{\epsilon+K^2}{\epsilon}\right) \cdot \exp\left(-\frac{r^2}{2\sigma^2} \cdot \frac{K^2+\epsilon}{K^2}\right)\right)^{\frac{\sigma^2}{K^2+\varepsilon}} \\
&= C \exp\left(-\frac{r^2}{2K^2}\right)
\end{aligned}$$

When $\rho_\sigma(r) \leq 1$, since $\frac{\sigma^2}{K^2+\varepsilon} < 1$ due to $\sigma < K$ and $\varepsilon > 0$, we have

$$\sqrt{2\pi}\sigma \cdot \rho_\sigma(r) \leq \sqrt{2\pi}\sigma \cdot \rho_\sigma(r)^{\frac{\sigma^2}{K^2+\varepsilon}}.$$

Therefore,

$$1 - F_\sigma(r) \leq (M_2 + \sqrt{2\pi}\sigma) \cdot \rho_\sigma(r)^{\frac{\sigma^2}{K^2+\varepsilon}}.$$

When $\rho_\sigma(r) > 1$, we will also have $\rho_\sigma(r)^{\frac{\sigma^2}{K^2+\varepsilon}} > 1$. Hence the following inequality holds

$$1 - F_\sigma(r) \leq 1 < \rho_\sigma(r)^{\frac{\sigma^2}{K^2+\varepsilon}}.$$

Above all, if we choose $M = \max\{M_1, M_2 + \sqrt{2\pi}\sigma, 1\}$, then we have

$$1 - F_\sigma(r) \leq M\rho_\sigma(r)^{\frac{\sigma^2}{K^2+\varepsilon}}, \quad \forall r \geq 0.$$

□

Proposition 11. *For two distribution $\mathbb{P}, \mathbb{Q}$ on $\mathbb{R}$ with PDF $\rho_p(x), \rho_q(x)$ and CDF $F_p(x), F_q(x)$. Suppose $\rho_p(x), \rho_q(x) > 0$ for every $x \in \mathbb{R}$, then*

$$W_2(\mathbb{P}, \mathbb{Q})^2 = \int_{-\infty}^{\infty} \rho_p(x) \left|F_q^{-1}(F_p(x)) - x\right|^2 dx,$$

where $F_q^{-1}(\cdot)$ is the inverse function of $F_q(\cdot)$.

Remark 5. *Here the map $x \to F_q^{-1}(F_p(x))$ is also the explicit form of the Brenier map $T_{p,q}$ between 1D distributions $\mathbb{P}$ and $\mathbb{Q}$. The Brenier map $x \to T_{p,q}(x)$ denotes the optimal coupling between $\mathbb{P}$ and $\mathbb{Q}$.*

Proof. According to [10], we have

$$W_2(\mathbb{P}, \mathbb{Q})^2 = \int_0^1 \left(F_q^{-1}(z) - F_p^{-1}(z)\right)^2 dz.$$

Since distribution $\mathbb{P}$ has PDF $\rho_p(x)$, we have $\frac{dF_p(x)}{dx} = \rho_p(x)$. We let $z = F_p(x)$ in the above equation. Then adopting changing of variables, we obtain that

$$W_2(\mathbb{P}, \mathbb{Q})^2 = \int_0^1 \left(F_q^{-1}(z) - F_p^{-1}(z)\right)^2 dz = \int_{-\infty}^{\infty} \rho_p(x) \left|F_q^{-1}(F_p(x)) - x\right|^2 dx.$$

□

Proposition 12. *Consider two 1D-distributions $\mathbb{P}, \mathbb{Q}$. Distribution $\mathbb{Q}$ is with always-positive PDF. We denote the PDF of $\mathbb{P}$ as $\rho_p(\cdot)$, and the CDFs of $\mathbb{P}, \mathbb{Q}$ as F_p, F_q*

respectively. If for some $\sigma > 0$ we have

$$\frac{\sup_{x\in[t-\sigma,t+\sigma]} |F_p(x) - F_q(x)|}{\inf_{x\in[t-\sigma,t+\sigma]} \rho_p(x)} \leq \sigma,$$

then

$$\left|F_q^{-1}(F_p(t)) - t\right| \leq \frac{\sup_{x\in[t-\sigma,t+\sigma]} |F_p(x) - F_q(x)|}{\inf_{x\in[t-\sigma,t+\sigma]} \rho_p(x)}.$$

To prove this proposition, we first provide a lemma.

Lemma 4. *Consider two 1D-distributions $\mathbb{P}, \mathbb{Q}$. Distribution $\mathbb{Q}$ is with always-positive PDF. Suppose the CDFs of $\mathbb{P}, \mathbb{Q}$ are F_p, F_q, and let $X \sim \mathbb{P}$. Then if*

$$\mathbf{P}(X \in (r-\alpha, r]) \geq |F_p(r-\alpha) - F_q(r-\alpha)|$$
$$\mathbf{P}(X \in (r, r+\alpha]) \geq |F_p(r+\alpha) - F_q(r+\alpha)|$$

both hold, we will have

$$|F_q^{-1}(F_p(r)) - r| \leq \delta,$$

where F_q^{-1} is the inverse function of F_q.

Proof. We notice that

$$\begin{aligned}
F_q(r-\alpha) &\leq F_p(r-\alpha) + |F_p(r-\alpha) - F_q(r-\alpha)| \\
&= F_p(r) - \mathbf{P}(X \in (r-\alpha, r]) + |F_p(r-\alpha) - F_q(r-\alpha)| \\
&\leq F_p(r).
\end{aligned}$$

Similarly, we also obtain that

$$F_p(r) \leq F_q(r+\alpha).$$

Therefore, $F_q(r-\alpha) \leq F_p(r) \leq F_q(r+\alpha)$.

Moreover, noticing that $\mathbb{Q}$ is with always-positive PDF, for any $\epsilon > 0$ we have

$$F_q(r-\delta-\epsilon) < F_q(r-\delta) \leq F_p(r) \leq F_q(r+\alpha) < F_q(r+\delta+\epsilon).$$

Since F_q and F_p are both non-decreasing functions, we have

$$|F_q^{-1}(F_p(r)) - r| \leq \delta + \epsilon.$$

Choosing $\epsilon \to 0$, we obtain that

$$|F_q^{-1}(F_p(r)) - r| \leq \delta.$$

□

Equipped with this lemma, we are ready to prove Proposition 12.

Proof of Proposition 12. We let random variable $X \sim \mathbb{P}$ and

$$\alpha = \frac{\sup_{x\in[t-\sigma,t+\sigma]} |F_p(x) - F_q(x)|}{\inf_{x\in[t-\sigma,t+\sigma]} \rho_p(x)}.$$

Then the assumption indicates that $0 < \alpha \leq \sigma$. Hence we obtain that

$$\begin{aligned}
\mathbf{P}(X \in (t-\alpha, t])) &\geq \alpha \cdot \inf_{x\in[t-\alpha,t+\alpha]} \rho_p(x) \geq \alpha \cdot \inf_{x\in[t-\sigma,t+\sigma]} \rho_p(x), \\
\mathbf{P}(X \in (t, t+\alpha])) &\geq \alpha \cdot \inf_{x\in[t-\alpha,t+\alpha]} \rho_p(x) \geq \alpha \cdot \inf_{x\in[t-\sigma,t+\sigma]} \rho_p(x).
\end{aligned}$$

Therefore,

$$\begin{aligned}
|F_p(t-\alpha) - F_q(t-\alpha)| &\leq \sup_{x\in[t-\sigma,t+\sigma]} |F_p(x) - F_q(x)| \\
&= \alpha \cdot \inf_{x\in[t-\sigma,t+\sigma]} \rho_p(x) \leq \mathbf{P}(X \in (t-\alpha, t])), \\
|F_p(t+\alpha) - F_q(t+\alpha)| &\leq \sup_{x\in[t-\sigma,t+\sigma]} |F_p(x) - F_q(x)| \\
&= \alpha \cdot \inf_{x\in[t-\sigma,t+\sigma]} \rho_p(x) \leq \mathbf{P}(X \in (t+\alpha, t]))
\end{aligned}$$

Therefore, according to Lemma 4, we obtain that

$$\left|F_q^{-1}(F_p(t)) - t\right| \leq \alpha = \frac{\sup_{x\in[t-\sigma,t+\sigma]} |F_p(x) - F_q(x)|}{\inf_{x\in[t-\sigma,t+\sigma]} \rho_p(x)}.$$

$\square$

Proposition 13. *Suppose distribution $\mathbb{P}, \mathbb{Q}$ satisfies that for any $r \geq 0$,*

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

where $X \sim \mathbb{P}, \quad Y \sim \mathbb{Q}$. We use $F_{p,\sigma}, F_{q,\sigma}$ to denote the CDFs of distribution $\mathbb{P} * \mathcal{N}(0,\sigma^2), \mathbb{Q} * \mathcal{N}(0,\sigma^2)$ separately. Then for any $R \geq 0$ and $x \in [-R, R]$ we have

$$\left|F_{q,\sigma}^{-1}(F_{p,\sigma}(x)) - x\right| \leq 2R + 2\sigma + \frac{K_2 R}{\sigma} + K_1\sqrt{2\log(2C_1)} + K_2\sqrt{2\log\left(\frac{4RC_2}{\sigma}\right)}.$$

Proof. First we notice that the PDFs of distribution $\mathbb{P} * \mathcal{N}(0,\sigma^2), \mathbb{Q} * \mathcal{N}(0,\sigma^2)$ at any real number is positive, hence $F_{p,\sigma}, F_{q,\sigma}$ are monotonically increasing in the entire real line. In the following, we use Φ_σ to denote the CDF of distribution $\mathcal{N}(0,\sigma^2)$. We have

$$\mathbf{P}\left(|X| \geq K_1\sqrt{2\log(2C_1)}\right) \leq C_1 \exp\left(-\log(2C_1)\right) = \frac{1}{2}.$$

Therefore, we obtain that

$$\mathbf{P}\left(|X| \leq K_1\sqrt{2\log(2C_1)}\right) \geq 1 - \frac{1}{2} = \frac{1}{2}.$$

We further notice that if $X \sim \mathbb{P}, Z \sim \mathcal{N}(0,\sigma^2)$ are independent, $X + Z \sim \mathbb{P} * \mathcal{N}(0,\sigma^2)$. And also for any $R \geq 0$,

$$\begin{gathered}\{|X| \leq K_1\sqrt{2\log(2C_1)}\} \cap \{Z \leq -K_1\sqrt{2\log(2C_1)} - R\} \subset \{X + Z \leq -R\} \\ \{|X| \leq K_1\sqrt{2\log(2C_1)}\} \cap \{Z \geq K_1\sqrt{2\log(2C_1)} + R\} \subset \{X + Z \geq R\}.\end{gathered}$$

Hence noticing that $\Phi_\sigma(-R - K_1\sqrt{2\log(2C_1)}) = 1 - \Phi_\sigma(R + K_1\sqrt{2\log(2C_1)}) =$

$\mathbf{P}(Z \leq -K_1\sqrt{2\log(2C_1)} - R) = \mathbb{P}(Z \geq K_1\sqrt{2\log(2C_1)} + R)$, we have

$$
\begin{aligned}
&\frac{1}{2}\Phi_\sigma(-R - K_1\sqrt{2\log(2C_1)}) \\
&\leq \mathbf{P}\left(|X| \geq K_1\sqrt{2\log(2C_1)}\right)\mathbf{P}(Z \leq -K_1\sqrt{2\log(2C_1)} - R) \\
&\leq \mathbf{P}(X + Z \leq -R) = F_{p,\sigma}(-R) \\
&\frac{1}{2}\Phi_\sigma(-R - K_1\sqrt{2\log(2C_1)}) \\
&\leq \mathbf{P}\left(|X| \geq K_1\sqrt{2\log(2C_1)}\right)\mathbf{P}(Z \geq K_1\sqrt{2\log(2C_1)} + R) \\
&\leq \mathbf{P}(X + Z \geq R) = 1 - F_{p,\sigma}(R),
\end{aligned}
$$

which indicates that

$$
\frac{1}{2}\Phi_\sigma(-R - K_1\sqrt{2\log(2C_1)}) \leq F_{p,\sigma}(-R) \leq F_{p,\sigma}(R) \leq 1 - \frac{1}{2}\Phi_\sigma(-R - K_1\sqrt{2\log(2C_1)})
$$

after noticing that $F_{p,\sigma}(-R) \leq F_{p,\sigma}(R)$ due to the monotonicity of $F_{p,\sigma}$.

Next, if $Y \sim \mathbb{Q}, Z \sim \mathcal{N}(0, \sigma^2)$ are independent, we have $Y + Z \sim \mathbb{Q} * \mathcal{N}(0, \sigma^2)$. Noticing that for $\forall R, q \geq 0$, we have

$$
\begin{aligned}
\{Y + Z \leq -R - q\} &\subset \{Z \leq -R\} \cup \{Y \leq -q\}, \\
\{Y + Z \geq R + q\} &\subset \{Z \geq R\} \cup \{Y \geq q\},
\end{aligned}
$$

we obtain that

$$
\begin{aligned}
F_{q,\sigma}(-R - q) &\leq \Phi_\sigma(-R) + \mathbf{P}(|Y| \geq q), \\
1 - F_{q,\sigma}(R + q) &\leq 1 - \Phi_\sigma(R) + \mathbf{P}(|Y| \geq q) = \Phi_\sigma(-R) + \mathbf{P}(|Y| \geq q).
\end{aligned}
$$

According to Proposition 2.1.2 in [15], we have

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

Hence when $R \geq 2\sigma$, we will have

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

We further notice that

$$\mathbf{P}(|Y| \geq q) \leq C_2 \exp\left(-\frac{q^2}{2K_2^2}\right).$$

Therefore, when

$$q \geq \frac{K_2 R}{\sigma} + K_2 \sqrt{2\log\left(\frac{4RC_2}{\sigma}\right)},$$

we would have

$$\mathbf{P}(|Y| \geq q) \leq C_2 \exp\left(-\frac{q^2}{2K_2^2}\right) \leq \frac{\sigma}{4R} \exp\left(-\frac{R^2}{2\sigma^2}\right) \leq \Phi_\sigma(-R),$$

which indicates that

$$F_{q,\sigma}(-R-q) \leq 2\Phi_\sigma(-R), \quad 1 - F_{q,\sigma}(R+q) \leq 2\Phi_\sigma(-R).$$

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

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

This indicates that

$$\begin{aligned}
&\frac{1}{4}\Phi_\sigma(-R - K_1\sqrt{2\log(2C_1)}) \\
&= \frac{1}{4} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{-R-K_1\sqrt{2\log(2C_1)}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \\
&\geq \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{-R-K_1\sqrt{2\log(2C_1)}} \exp\left(-\frac{(x-2\sigma)^2}{2\sigma^2}\right) dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{-R-K_1\sqrt{2\log(2C_1)}-2\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \\
&= \Phi_\sigma(-R - K_1\sqrt{2\log(2C_1)} - 2\sigma).
\end{aligned}$$

Therefore, we obtain that

$$
\begin{aligned}
&F_{q,\sigma}\left(-R - K_1\sqrt{2\log(2C_1)} - 2\sigma - \frac{K_2 R}{\sigma} - K_2\sqrt{2\log\left(\frac{4RC_2}{\sigma}\right)}\right) \\
&\leq 2\Phi_\sigma\left(-R - K_1\sqrt{2\log(2C_1)} - 2\sigma\right) \leq \frac{1}{2}\Phi_\sigma(-R - K_1\sqrt{2\log(2C_1)}) \leq F_{p,\sigma}(-R).
\end{aligned}
$$

Similarly, we can also obtain that

$$
1 - F_{q,\sigma}\left(R + K_1\sqrt{2\log(2C_1)} + 2\sigma + \frac{K_2 R}{\sigma} + K_2\sqrt{2\log\left(\frac{4RC_2}{\sigma}\right)}\right) \leq 1 - F_{p,\sigma}(R).
$$

Hence using the monotonicity of $F_{p,\sigma}$ and $F_{q,\sigma}$, we obtain that for any $R \geq 0$ and $x \in [-R, R]$,

$$
\begin{aligned}
&F_{q,\sigma}\left(-R - K_1\sqrt{2\log(2C_1)} - 2\sigma - \frac{K_2 R}{\sigma} - K_2\sqrt{2\log\left(\frac{4RC_2}{\sigma}\right)}\right) \\
&\leq F_{p,\sigma}(x) \leq F_{q,\sigma}\left(R + K_1\sqrt{2\log(2C_1)} + 2\sigma + \frac{K_2 R}{\sigma} + K_2\sqrt{2\log\left(\frac{4RC_2}{\sigma}\right)}\right),
\end{aligned}
$$

which indicates that

$$
\begin{aligned}
&- R - K_1\sqrt{2\log(2C_1)} - 2\sigma - \frac{K_2 R}{\sigma} - K_2\sqrt{2\log\left(\frac{4RC_2}{\sigma}\right)} \\
&\leq F_{q,\sigma}^{-1}(F_{p,\sigma}(x)) \leq R + K_1\sqrt{2\log(2C_1)} + 2\sigma + \frac{K_2 R}{\sigma} + K_2\sqrt{2\log\left(\frac{4RC_2}{\sigma}\right)}.
\end{aligned}
$$

Hence we have

$$
\left|F_{q,\sigma}^{-1}(F_{p,\sigma}(x)) - x\right| \leq 2R + K_1\sqrt{2\log(2C_1)} + 2\sigma + \frac{K_2 R}{\sigma} + K_2\sqrt{2\log\left(\frac{4RC_2}{\sigma}\right)}.
$$

$\square$

Proposition 14. *Suppose $F_\sigma, \tilde{F}_{n,\sigma}$ are CDFs of distribution $\mathbb{P} * \mathcal{N}(0, \sigma^2)$ and $\mathbb{P}_n *$*

$\mathcal{N}(0, \sigma^2)$ *respectively. We define*

$$G(t) = \begin{cases} \frac{1}{n} & 0 \le t \le \frac{1}{n} \quad \frac{n-1}{n} \le t \le 1, \\ t & \frac{1}{n} \le t \le \frac{1}{2}, \\ 1 - t & \frac{1}{2} \le t \le \frac{n-1}{n}. \end{cases}$$

Then with probability at least $1 - \delta$, *we have the following inequality:*

$$\sup_{t \in \mathbb{R}} \frac{|F_\sigma(t) - \tilde{F}_{n,\sigma}(t)|}{\sqrt{G(F(t))}} \le \frac{16}{\sqrt{n}} \log\left(\frac{2n}{\delta}\right).$$

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

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

$$\sup_{t \in \mathbb{R}} \frac{|F(t) - \hat{F}(t)|}{\sqrt{G(F(t))}} \le 8\sqrt{\frac{1}{n}} \log\left(\frac{n}{\delta}\right).$$

Remark 6. *This lemma follows directly from Theorem 2.1 of [6], if we choose* $r = \frac{1}{n}, \delta = \frac{1}{2}, s_j = K\log\left(\frac{2K}{\delta'}\right)$.

Remark 7. *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*

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

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*

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

Proof of Lemma 5. Since $\mathbb{P}$ has positive PDF at every points on $\mathbb{R}$, its CDF F has inverse function F^{-1}. We consider the distribution $\mathbb{Q}$ on $[0,1]$ such that $\mathbb{Q}(\cdot) = \mathbb{P}(F^{-1}(\cdot))$. Then $\mathbb{Q}$ matches the uniform distribution on $[0,1]$. We let $\mathbb{Q}_n(\cdot) = \mathbb{P}_n(F^{-1}(\cdot))$. Then the distribution of $\mathbb{P}_n$ is equivalent to the distribution of n-point empirical measure of $\mathbb{Q}$. Therefore, we only need to prove this lemma under the assumption that $\mathbb{Q}$ is the uniform distribution on $[0,1]$.

In the next, we assume $\mathbb{P}$ to be the uniform distribution $[0,1]$, and will have $F(t)=t$ for any $0\leq t\leq 1$. We only need to prove that with probability at least $1-\delta$,

$$\sup_{t\in\mathbb{R}} \frac{|F(t)-\hat{F}(t)|}{\sqrt{G(t)}} \leq \sqrt{\frac{\log n}{n}}.$$

According to Bernstein inequality, for any $1\leq k\leq \frac{n}{2}$, if we let $t=\frac{k}{n}$, then we have

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

Therefore, applying union bound for $1\leq k\leq \frac{n}{2}$, we obtain that

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

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

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

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 $\left|F\left(\frac{k}{n}\right)-\hat{F}\left(\frac{k}{n}\right)\right| \leq 4\sqrt{\frac{(k/n)}{n}}\log\left(\frac{n}{\delta}\right)$, then for every $0\leq t\leq\frac{1}{n}$, we have

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

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

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

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

$$\frac{|F(t)-\hat{F}(t)|}{\sqrt{G(t)}} \le 8\sqrt{\frac{1}{n}}\log\left(\frac{n}{\delta}\right)$$

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

$$\sup_{0\le t\le 1}\frac{|F(t)-\hat{F}(t)|}{\sqrt{G(t)}} \le 8\sqrt{\frac{1}{n}}\log\left(\frac{n}{\delta}\right).$$

This completes the proof of this lemma.

$\square$

Proof of Proposition 14. Suppose random variables $X \sim \mathbb{P}, Y \sim \mathcal{N}(0,\sigma^2)$ are independent. Then $X+Y \sim \mathbb{P} * \mathcal{N}(0,\sigma^2)$. 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 $\mathbb{P} * \mathcal{N}(0,\sigma^2)$. We use $\hat{F}$ to denote the PDF of empirical measure $\hat{\mathbb{P}} = \frac{1}{n}\sum_{i=1}^n \delta_{X_i+Y_i}$. Then according to Lemma 5, we have with probability $1-\delta$,

$$\sup_{t\in\mathbb{R}}\frac{|F(t)-\hat{F}(t)|}{\sqrt{G(t)}} \le 8\sqrt{\frac{1}{n}}\log\left(\frac{n}{\delta}\right).$$

Hence Markov inequality indicates that

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

Therefore, we have

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

Moreover, we notice that

$$
\mathbb{E}\left[\hat{F}(t)\middle|X_1,\cdots,X_n\right]=\mathbf{P}\left(\frac{1}{n}\sum_{i=1}^n(X_i+Y_i)\leq t\middle|X_1,\cdots,X_n\right)=\tilde{F}_{n,\sigma}(t),
$$

where $\tilde{F}_{n,\sigma}$ is the CDF of $\mathbb{P}_n*\mathcal{N}(0,\sigma^2)$ 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

$$
\begin{aligned}
&\mathbb{E}\left[\exp\left(\sup_{t\in\mathbb{R}}\frac{\sqrt{n}}{16}\cdot\frac{|F(t)-\tilde{F}_{n,\sigma}(t)|}{\sqrt{G(t)}}-\log n\right)\right]\\
&\leq\mathbb{E}\left[\exp\left(\sup_{t\in\mathbb{R}}\frac{\sqrt{n}}{16}\cdot\frac{|F(t)-\tilde{F}_{n,\sigma}(t)|}{\sqrt{G(t)}}-\frac{\log n}{2}\right)\right]\\
&=\mathbb{E}\left[\exp\left(\sup_{t\in\mathbb{R}}\frac{\sqrt{n}}{16}\cdot\frac{|F(t)-\mathbb{E}_{Y_i,1\leq i\leq n}[\hat{F}(t)]|}{\sqrt{G(t)}}-\frac{\log n}{2}\right)\right]\\
&\leq\mathbb{E}\left[\exp\left(\sup_{t\in\mathbb{R}}\frac{\sqrt{n}}{16}\cdot\frac{\mathbb{E}_{Y_i,1\leq i\leq n}|F(t)-\hat{F}(t)|}{\sqrt{G(t)}}-\frac{\log n}{2}\right)\right]\\
&\leq\mathbb{E}\left[\exp\left(\mathbb{E}\left[\sup_{t\in\mathbb{R}}\frac{\sqrt{n}}{16}\cdot\frac{|F(t)-\hat{F}(t)|}{\sqrt{G(t)}}-\frac{\log n}{2}\middle|X_1,\cdots,X_n\right]\right)\right]\\
&\leq\mathbb{E}\left[\mathbb{E}\left[\exp\left(\sup_{t\in\mathbb{R}}\frac{\sqrt{n}}{16}\cdot\frac{|F(t)-\hat{F}(t)|}{\sqrt{G(t)}}-\frac{\log n}{2}\right)\right]\middle|X_1,\cdots,X_n\right]\\
&\leq 2.
\end{aligned}
$$

And according to Markov inequality, we have

$$\mathbf{P}\left(\exp\left(\sup_{t\in\mathbb{R}}\frac{\sqrt{n}}{16}\cdot\frac{|F(t)-\tilde{F}_{n,\sigma}(t)|}{\sqrt{G(t)}}-\log n\right)\geq\frac{2}{\delta}\right)\leq\delta.$$

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

$$\sup_{t\in\mathbb{R}}\frac{|\tilde{F}_{n,\sigma}(t)-F(t)|}{\sqrt{G(t)}}\leq\frac{16}{\sqrt{n}}\log\left(\frac{2n}{\delta}\right).$$

$\square$

Proposition 15. *Suppose distribution $\mathbb{P}$ is a K-subgaussian distribution, e.g.*

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

and distribution $\mathbb{P}_n$ is the empirical distribution obtained through n i.i.d. samples from $\mathbb{P}$. We suppose $F_\sigma,\tilde{F}_{n,\sigma}$ to be the CDFs of distribution $\mathbb{P}*\mathcal{N}(0,\sigma^2),\mathbb{P}_n*\mathcal{N}(0,\sigma^2)$. Then for any $0<\delta<1$, with probability at least $1-\delta$, we have for any $R\geq 0$ and $x\in[-R,R]$,

$$\left|\tilde{F}_{n,\sigma}^{-1}(F_\sigma(x))-x\right|\leq 2R+2\sigma+\frac{KR}{\sigma}+K\sqrt{2\log(2C)}+K\sqrt{2\log\left(\frac{4CRn}{\delta\sigma}\right)}.$$

Proof. Since $\mathbf{P}(|X|\geq r)\leq C\exp\left(-\frac{r^2}{2K^2}\right)$, with probability at least

$$\left(1-C\exp\left(-\frac{r^2}{2K^2}\right)\right)^n\geq 1-nC\exp\left(-\frac{r^2}{2K^2}\right)$$

we have $\mathrm{supp}(\mathbb{P}_n)\subset[-r,r]$. Choosing

$$r=K\sqrt{2\log\left(\frac{Cn}{\delta}\right)},$$

we obtain that with probability at least $1 - \delta$,

$$\operatorname{supp}(\mathbb{P}_n) \subset [-r, r] = \left[-K\sqrt{2\log\left(\frac{Cn}{\delta}\right)}, K\sqrt{2\log\left(\frac{Cn}{\delta}\right)} \right].$$

Therefore, choosing $K_2 = K$ and $C_2 = \frac{Cn}{\delta}$ and letting $Y \sim \mathbb{P}_n$, with probability at least $1 - \delta$ we have

$$\mathbf{P}(|Y| \geq q) \leq C_2 \exp\left(-\frac{q^2}{2K^2}\right), \quad \forall q \geq 0.$$

Adopting Proposition 13 with

$$\mathbb{P} = \mathbb{P}, \mathbb{Q} = \mathbb{P}_n, K_1 = K_2 = K, C_1 = C, C_2 = \frac{Cn}{\delta},$$

we obtain that with probability at least $1 - \delta$, for any $R \geq 0$ and $x \in [-R, R]$, we have

$$\left| \tilde{F}_{n,\sigma}^{-1}(F_\sigma(x)) - x \right| \leq 2R + 2\sigma + \frac{KR}{\sigma} + K\sqrt{2\log(2C)} + K\sqrt{2\log\left(\frac{4CRn}{\delta\sigma}\right)}.$$

□

Equipped with these propositions, we are ready to prove the upper bound part of Theorem 2.

Proof of Upper Bound Part of Theorem 2. We use ρ_σ to denote the PDF of $\mathbb{P} * \mathcal{N}(0, \sigma^2)$, and use $F_\sigma, \tilde{F}_{n,\sigma}$ to denote the CDF of $\mathbb{P} * \mathcal{N}(0, \sigma^2)$ and $\mathbb{P}_n * \mathcal{N}(0, \sigma^2)$ respectively. Then according to the Proposition 14, we have

$$\sup_{t \leq 0} \frac{|F_\sigma(t) - \tilde{F}_{n,\sigma}(t)|}{\sqrt{G(t)}} \leq \frac{16}{\sqrt{n}} \log\left(2n^2\right)$$

holds with probability at least $1 - 1/n$. In the main part of the proof, we assume that this event holds, and scenarios where this event does not hold will be discussed at the end of this proof.

According to Proposition 10, for any positive $\varepsilon > 0$ (the value of ε will be specified later), we have

$$\begin{cases} F_\sigma(t) \leq M\rho_\sigma(t)^{\frac{\sigma^2}{K^2+\varepsilon}}, & t \leq 0, \\ 1 - F_\sigma(t) \leq M\rho_\sigma(t)^{\frac{\sigma^2}{K^2+\varepsilon}}, & t > 0, \end{cases}$$

where we write $M = M(\varepsilon, \sigma, K, C)$ in Proposition 10. Therefore noticing that

$$G(t) = \min\{\max\{t, 1/n\}, \max\{1-t, 1/n\}\},$$

we have $\forall t \in \mathbb{R}$,

$$|F_\sigma(t) - \tilde{F}_{n,\sigma}(t)| \leq \max\left\{\frac{16}{n}\log\left(2n^2\right), \frac{16\sqrt{M}}{\sqrt{n}}\log\left(2n^2\right)\rho_\sigma(t)^{\frac{\sigma^2}{2K^2+2\varepsilon}}\right\}.$$

In the following, we first assume the above inequality holds. In order to analyze the behavior of $|F_\sigma(x) - \tilde{F}_{n,\sigma}(x)|$ for $x \in [t-\sigma, t+\sigma]$, we let

$$L_n(t) = \max\left\{\frac{16}{n}\log\left(2n^2\right), \sup_{t-\sigma \leq x \leq t+\sigma}\frac{16\sqrt{M}}{\sqrt{n}}\log\left(2n^2\right)\rho_\sigma(x)^{\frac{\sigma^2}{2K^2+2\varepsilon}}\right\},$$

then $\forall t \in \mathbb{R}, x \in [t-\sigma, t+\sigma]$, we have

$$|F_\sigma(x) - \tilde{F}_{n,\sigma}(x)| \leq L_n(t). \tag{5.2}$$

According to Proposition 7 we have $0 < \rho_\sigma(t) \leq \frac{1}{\sqrt{2\pi}\sigma}$ for every $t \in \mathbb{R}$. Hence leting

$$a(t) = \sqrt{2\sigma^2 \log\frac{1}{\sqrt{2\pi}\sigma\rho_\sigma(t)}},$$

we will have

$$a(t) \in [0, \infty), \quad \rho_\sigma(t) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{a(t)^2}{2\sigma^2}\right), \qquad \forall t \in \mathbb{R}.$$

Based on the value of $a(t)$ for $t \in \mathbb{R}$, we divide $\mathbb{R}$ into the following three nonintersected measurable sets:

$$
\begin{aligned}
A = \ & \{t | a(t) \leq 5\sigma\}, \\
B = & \left\{ t \middle| 5\sigma < a(t) \leq \Delta_n \triangleq \min\left\{ \sqrt{2\sigma^2} \cdot \sqrt{\log n - \log\left(16\sqrt{2\pi}\log(2n^2)\right)} - 5\sigma, \right.\right. \\
& \left.\left. \sqrt{\left(\frac{1}{2\sigma^2} - \frac{1}{4K^2+4\varepsilon}\right)^{-1} \log\left(\frac{\sigma\sqrt{n}}{\sqrt{2\pi}\sigma M'}\log(2n^2)^{-1}\exp\left(-\frac{25}{2}\right)\right)} - \frac{\sigma^2+2K^2+2\varepsilon}{2K^2+2\varepsilon-\sigma^2}\cdot(5\sigma) \right\} \right. \\
C = & \{t | a(t) > \Delta_n\}.
\end{aligned}
$$

Then $\mathbb{R} = A \cup B \cup C$.

We let

$$
\underline{\rho_\sigma}(t) = \inf_{t-\sigma \leq x \leq t+\sigma} \rho_\sigma(x), \quad \overline{\rho_\sigma}(t) = \sup_{t-\sigma \leq x \leq t+\sigma} \rho_\sigma(x).
$$

According to Proposition 8, for every $t - \sigma \leq x \leq t + \sigma$ we have

$$
\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(a(t)+5\sigma)^2}{2\sigma^2}\right) \leq \rho_\sigma(x) \leq \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\max\{0, a(t)-5\sigma\}^2}{2\sigma^2}\right).
$$

Hence we have

$$
\underline{\rho_\sigma}(t) \geq \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(a(t)+5\sigma)^2}{2\sigma^2}\right), \quad \overline{\rho_\sigma}(t) \leq \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\max\{0, a(t)-5\sigma\}^2}{2\sigma^2}\right).
$$

Therefore, we can obtain an upper bound on $L_n(t)$:

$$
\begin{aligned}
L_n(t) &= \max\left\{ \frac{16}{n}\log\left(2n^2\right), \frac{16\sqrt{M}}{\sqrt{n}}\log\left(2n^2\right)\overline{\rho_\sigma}(x)^{\frac{\sigma^2}{2K^2+2\varepsilon}} \right\} \\
&\leq \max\left\{ \frac{16}{n}\log\left(2n^2\right), \frac{16\sqrt{M}}{\sqrt{n}}\log\left(2n^2\right)\left(\sqrt{2\pi}\sigma\right)^{-\frac{\sigma^2}{2K^2+2\varepsilon}}\exp\left(-\frac{\max\{0, a(t)-5\sigma\}^2}{4K^2+4\varepsilon}\right) \right\}.
\end{aligned}
$$

If we let $M' = 16 + 16\sqrt{M} \cdot \left(\sqrt{2\pi}\sigma\right)^{\frac{\sigma^2}{2K^2+2\varepsilon}} > 16$, then we will have

$$
L_n(t) \leq \max\left\{ \frac{16}{n}\log\left(2n^2\right), \frac{M'}{\sqrt{n}}\log\left(2n^2\right)\exp\left(-\frac{\max\{0, a(t)-5\sigma\}^2}{4K^2+4\varepsilon}\right) \right\}
$$

For $t \in A$, we have $a(t) \le 5\sigma$. Since

$$\exp\left(-\frac{\max\{0, a(t) - 5\sigma\}^2}{4K^2 + 4\varepsilon}\right) \le 1, \quad \frac{16}{n} \le \frac{16}{\sqrt{n}} \le \frac{M'}{\sqrt{n}},$$

we have

$$L_n(t) \le \max\left\{\frac{16}{n}\log(2n^2), \frac{M'}{\sqrt{n}}\log(2n^2)\right\} \le \frac{M'}{\sqrt{n}}\log\left(2n^2\right).$$

Moreover, we have

$$\underline{\rho_\sigma}(t) \ge \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(a(t)+5\sigma)^2}{2\sigma^2}\right) \ge \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(10\sigma)^2}{2\sigma^2}\right) = \frac{\exp(-50)}{\sqrt{2\pi}\sigma}.$$

When n is large enough such that

$$\frac{M'}{\sqrt{n}}\log\left(2n^2\right) \le \frac{\exp(-50)}{\sqrt{2\pi}} = \sigma \cdot \frac{\exp(-50)}{\sqrt{2\pi}\sigma} \le \sigma \cdot \underline{\rho_\sigma}(t),$$

$\forall t \in A$ we have

$$\frac{\sup_{x\in[t-\sigma,t+\sigma]}|F_\sigma(x) - \tilde{F}_{n,\sigma}(x)|}{\inf_{x\in[t-\sigma,t+\sigma]}\rho_\sigma(x)} \le \frac{L_n(t)}{\underline{\rho_\sigma}(t)} \le \sigma.$$

We further notice that distribution $\mathbb{P}, \mathbb{P}_n$ are both with always-positive PDFs (Proposition 7). Hence Proposition 12 indicates that

$$\left|\tilde{F}_{n,\sigma}^{-1}(F_\sigma(t)) - t\right| \le \frac{\sup_{x\in[t-\sigma,t+\sigma]}|F_\sigma(x) - \tilde{F}_{n,\sigma}(x)|}{\inf_{x\in[t-\sigma,t+\sigma]}\rho_\sigma(x)} \le \frac{L_n(t)}{\underline{\rho_\sigma}(t)} \le \frac{M'\sqrt{2\pi}\sigma}{\sqrt{n}\exp(-50)}\log\left(2n^2\right).$$

This indicates that

$$\begin{aligned}
&\int_A \rho_\sigma(t)\left|\tilde{F}_{n,\sigma}^{-1}(F_\sigma(t)) - t\right|^2 dt \\
\le &\int_A \rho_\sigma(t)dt \cdot \left(\frac{16\sqrt{2\pi}\sigma}{\sqrt{n}\exp(-50)}(1+C)\log\left(2n^2\right)\right)^2 \\
= &\tilde{\mathcal{O}}\left(\frac{1}{n}\right).
\end{aligned}$$

For $t \in B$, we have $a(t) \geq 5\sigma$, which indicates that

$$L_n(t) \leq \max\left\{\frac{16}{n}\log\left(2n^2\right), \frac{M'}{\sqrt{n}}\log\left(2n^2\right)\exp\left(-\frac{(a(t)-5\sigma)^2}{4K^2+4\varepsilon}\right)\right\}.$$

Noticing that

$$\begin{aligned}
a(t) \leq \Delta_n \leq & \sqrt{\left(\frac{1}{2\sigma^2}-\frac{1}{4K^2+4\varepsilon}\right)^{-1}\log\left(\frac{\sigma\sqrt{n}}{\sqrt{2\pi}\sigma M'}\log(2n^2)^{-1}\exp\left(-\frac{25}{2}\right)\right)} \\
& -\frac{\sigma^2+2K^2+2\varepsilon}{2K^2+2\varepsilon-\sigma^2}\cdot(5\sigma),
\end{aligned}$$

we obtain

$$\begin{aligned}
& \exp\left(-\frac{(a(t)-5\sigma)^2}{4K^2+4\varepsilon}+\frac{(a(t)+5\sigma)^2}{2\sigma^2}\right) \\
= & \exp\left(-\frac{a(t)^2}{4K^2+4\varepsilon}+\frac{5a(t)\sigma}{2K^2+2\varepsilon}-\frac{25\sigma^2}{4K^2+4\varepsilon}+\frac{a(t)^2}{2\sigma^2}+\frac{5a(t)\sigma}{\sigma^2}+\frac{25\sigma^2}{2\sigma^2}\right) \\
\leq & \exp\left(\left(\frac{1}{2\sigma^2}-\frac{1}{4K^2+4\varepsilon}\right)a(t)^2+\left(\frac{1}{2K^2+2\varepsilon}+\frac{1}{\sigma^2}\right)\cdot(5a(t)\sigma)+\frac{25}{2}\right) \\
\leq & \exp\left(\frac{25}{2}\right)\cdot\exp\left(\left(\frac{1}{2\sigma^2}-\frac{1}{4K^2+4\varepsilon}\right)\left(a(t)+\frac{\sigma^2+2K^2+2\varepsilon}{2K^2+2\varepsilon-\sigma^2}\cdot(5\sigma)\right)^2\right) \\
\leq & \frac{\sigma\sqrt{n}}{\sqrt{2\pi}\sigma M'}\log(2n^2)^{-1}.
\end{aligned}$$

Hence we have

$$\frac{M'}{\sqrt{n}}\cdot\exp\left(-\frac{(a(t)-5\sigma)^2}{4K^2+4\varepsilon}\right)\log\left(2n^2\right) \leq \sigma\cdot\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(a(t)+5\sigma)^2}{2\sigma^2}\right).$$

Moreover, according to the definition of the set B, we also notice that

$$\sigma\cdot\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(a(t)+5\sigma)^2}{2\sigma^2}\right) \geq \frac{16}{n}\log\left(2n^2\right).$$

Therefore, we obtain that

$$\begin{aligned}
&\sigma \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(a(t)+5\sigma)^2}{2\sigma^2}\right) \\
&\geq \max\left\{\frac{16}{n}\log\left(2n^2\right), \frac{M'}{\sqrt{n}}\log\left(2n^2\right)\exp\left(-\frac{(a(t)-5\sigma)^2}{4K^2+4\varepsilon}\right)\right\} \\
&= L_n(t)
\end{aligned}$$

Moreover, according to Proposition 8 we have

$$\underline{\rho_\sigma}(t) \geq \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(a(t)+5\sigma)^2}{2\sigma^2}\right),$$

which indicates that

$$\frac{\sup_{x\in[t-\sigma,t+\sigma]}|F_\sigma(x)-\tilde{F}_{n,\sigma}(x)|}{\inf_{x\in[t-\sigma,t+\sigma]}\rho_\sigma(x)} \leq \frac{L_n(t)}{\underline{\rho_\sigma}(t)} \leq \sigma.$$

Therefore, according to Proposition 12, we have

$$\left|\tilde{F}_{n,\sigma}^{-1}(F_\sigma(t))-t\right| \leq \frac{\sup_{x\in[t-\sigma,t+\sigma]}|F_\sigma(x)-\tilde{F}_{n,\sigma}(x)|}{\inf_{x\in[t-\sigma,t+\sigma]}\rho_\sigma(x)} \leq \frac{L(t)}{\underline{\rho_\sigma}(t)}.$$

According to our choice of a, we have for $\forall t \in B$,

$$\underline{\rho_\sigma}(t) \geq \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(a(t)+5\sigma)^2}{2\sigma^2}\right) = \tilde{\Omega}\left(n^{-\frac{K^2+\varepsilon}{2K^2+2\varepsilon-\sigma^2}}\right)$$

and also

$$L_n(t) = \tilde{\mathcal{O}}\left(n^{-\frac{K^2+\varepsilon}{2K^2+2\varepsilon-\sigma^2}}\right).$$

We further notice that for any $\epsilon > 0$,

$$\frac{\rho_\sigma(t)}{\underline{\rho_\sigma}(t)} \leq \frac{\exp\left(-\frac{a(t)^2}{2\sigma^2}\right)}{\exp\left(-\frac{(a(t)+5\sigma)^2}{2\sigma^2}\right)} = \exp\left(\frac{5a(t)}{\sigma}+\frac{25}{2}\right) = \tilde{\mathcal{O}}\left(n^\epsilon\right).$$

Therefore, we have for $\forall t \in B$,

$$\rho_\sigma(t) \cdot \left|\tilde{F}_{n,\sigma}^{-1}(F_\sigma(t)) - t\right|^2 \le \underline{\rho_\sigma}(t) \cdot \frac{\rho_\sigma(t)}{\underline{\rho_\sigma}(t)} \cdot \left(\frac{L(t)}{\underline{\rho_\sigma}(t)}\right)^2 = \frac{\rho_\sigma(t)}{\underline{\rho_\sigma}(t)} \cdot \frac{L(t)^2}{\underline{\rho_\sigma}(t)} = \tilde{\mathcal{O}}\left(n^{-\frac{K^2+\varepsilon}{2K^2+2\varepsilon-\sigma^2}+\epsilon}\right).$$

Additionally, we notice that for any $t \in B$, we have $a(t) \le \Delta$, hence for any $\epsilon > 0$,

$$\rho_\sigma(t) = \Omega\left(n^{-\frac{K^2+\varepsilon}{2K^2+2\varepsilon-\sigma^2}+\epsilon}\right).$$

According to Proposition 9, we have

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

This together with the above lower bound on $\rho_\sigma(t)$ provides a uniform upper bound Λ for $\forall t \in B$:

$$|t| \le \Lambda,$$

where we have

$$\Lambda = \mathcal{O}\left(2\sqrt{2}K\sqrt{\log\left(\left(C + \frac{1}{\sqrt{2\pi}\sigma}\right) n^{\frac{K^2+\varepsilon}{2K^2+2\varepsilon-\sigma^2}-\epsilon}\right)}\right) = \tilde{\mathcal{O}}(1).$$

Hence, $B \subset [-\Lambda, \Lambda]$. Therefore, with probability at least $1 - 1/n$, for any $\epsilon > 0$, we have

$$\begin{aligned}
&\int_B \rho_\sigma(t) \left|\tilde{F}_{n,\sigma}^{-1}(F_\sigma(t)) - t\right|^2 dt \\
&\le \int_{-\Delta_1}^{\Delta_1} \rho_\sigma(t) \left|\tilde{F}_{n,\sigma}^{-1}(F_\sigma(t)) - t\right|^2 dt \\
&\le 2\Delta \cdot \tilde{\mathcal{O}}\left(n^{-\frac{K^2+\varepsilon}{2K^2+2\varepsilon-\sigma^2}+\epsilon}\right) \\
&= \tilde{\mathcal{O}}\left(n^{-\frac{K^2+\varepsilon}{2K^2+2\varepsilon-\sigma^2}+\epsilon}\right)
\end{aligned}$$

Finally we consider $t \in C$, and hence $a(t) \ge \Delta$, which indicates that for any $\epsilon > 0$,

$$\rho_\sigma(t) = \tilde{\mathcal{O}}\left(n^{-\frac{K^2+\varepsilon}{2K^2+2\varepsilon-\sigma^2}+\epsilon}\right).$$

We define

$$R_0 \triangleq 2\sqrt{2}K\sqrt{\log\left(n\left(C+\frac{1}{\sqrt{2\pi}\sigma}\right)\right)} = \tilde{\mathcal{O}}(1).$$

Then for $X \sim \mathbb{P} * \mathcal{N}(0, \sigma^2)$, we have

$$\mathbf{P}(|X| \le R_0, X \in C) \le 2R_0 \cdot \tilde{\mathcal{O}}\left(n^{-\frac{K^2+\varepsilon}{2K^2+2\varepsilon-\sigma^2}+\epsilon}\right) = \tilde{\mathcal{O}}\left(n^{-\frac{K^2+\varepsilon}{2K^2+2\varepsilon-\sigma^2}+\epsilon}\right).$$

Moreover, according to Proposition 9, we have for any $l \in \mathbb{Z}_+$,

$$\mathbf{P}(|X| \ge lR_0) \le \left(C + \frac{1}{\sqrt{2\pi}\sigma}\right)\exp\left(-\frac{l^2R_0^2}{8K^2}\right) \le \frac{1}{n^{l^2}}.$$

And according to Proposition 15, with probability at least $1-\frac{1}{n}$, we have for any $R \ge 0$ and $t \in [-R, R]$,

$$\left|\tilde{F}_{n,\sigma}^{-1}(F_\sigma(t)) - t\right| \le 2R + 2\sigma + \frac{KR}{\sigma} + K\sqrt{2\log(2C)} + K\sqrt{2\log\left(\frac{4CRn^2}{\sigma}\right)}.$$

Hence, we obtain the following upper bound on the integral over C: with probability at least $1-1/n$, for any $\epsilon > 0$ we have

$$\begin{aligned}
&\int_C \rho_\sigma(t)\left|\tilde{F}_{n,\sigma}^{-1}(F_\sigma(t)) - t\right|^2 dt \\
&= \int_{C\cap\{|t|\le R_0\}} \rho_\sigma(t)\left|\tilde{F}_{n,\sigma}^{-1}(F_\sigma(t)) - t\right|^2 dt + \sum_{l=1}^{\infty}\int_{C\cap\{lR_0\le|t|\le(l+1)R_0\}} \rho_\sigma(t)\left|\tilde{F}_{n,\sigma}^{-1}(F_\sigma(t)) - t\right|^2 dt \\
&\le \mathbf{P}(|X| \le R_0, X\in C)\cdot\left(2R_0 + 2\sigma + \frac{KR_0}{\sigma} + K\sqrt{2\log(2C)} + K\sqrt{2\log\left(\frac{4CR_0n^2}{\sigma}\right)}\right)^2 \\
&\quad + \sum_{l=2}^{\infty}\mathbf{P}(|X| \ge lR_0, X\in C)\cdot\left(2lR_0 + 2\sigma + \frac{KlR_0}{\sigma} + K\sqrt{2\log(2C)} + K\sqrt{2\log\left(\frac{4ClR_0n^2}{\sigma}\right)}\right)^2 \\
&= \tilde{\mathcal{O}}\left(n^{-\frac{K^2+\varepsilon}{2K^2+2\varepsilon-\sigma^2}+\epsilon}\right)\cdot\tilde{\mathcal{O}}(1)^2 + \sum_{l=1}^{\infty} n^{-l^2}\cdot l^2\cdot\tilde{\mathcal{O}}(1)^2 \\
&= \tilde{\mathcal{O}}\left(n^{-\frac{K^2+\varepsilon}{2K^2+2\varepsilon-\sigma^2}+\epsilon}\right) + \tilde{\mathcal{O}}\left(n^{-1}\right) = \tilde{\mathcal{O}}\left(n^{-\frac{K^2+\varepsilon}{2K^2+2\varepsilon-\sigma^2}+\epsilon}\right).
\end{aligned}$$

Therefore, according to Proposition 11, combining these upper bounds on the integral over A, B, C together, we obtain that with probability at least $1-2/n$, for any $\epsilon > 0$,

$$
\begin{aligned}
W_2(\mathbb{P} * \mathcal{N}(0,\sigma^2), \mathbb{P}_n * \mathcal{N}(0,\sigma^2))^2 &= \int_{-\infty}^{\infty} \rho_\sigma(t) \left| \tilde{F}_{n,\sigma}^{-1}(F_\sigma(t)) - t \right|^2 dt, \\
= \int_A \rho_\sigma(t) \left| \tilde{F}_{n,\sigma}^{-1}(F_\sigma(t)) - t \right|^2 dt &+ \int_B \rho_\sigma(t) \left| \tilde{F}_{n,\sigma}^{-1}(F_\sigma(t)) - t \right|^2 dt \\
+ \int_C \rho_\sigma(t) \left| \tilde{F}_{n,\sigma}^{-1}(F_\sigma(t)) - t \right|^2 dt & \\
= \tilde{\mathcal{O}}\left(\frac{1}{n}\right) + \tilde{\mathcal{O}}\left(n^{-\frac{K^2+\varepsilon}{2K^2+2\varepsilon-\sigma^2}+\epsilon}\right) &+ \tilde{\mathcal{O}}\left(n^{-\frac{K^2+\varepsilon}{2K^2+2\varepsilon-\sigma^2}+\epsilon}\right) \\
= \tilde{\mathcal{O}}\left(n^{-\frac{K^2}{2K^2+2\varepsilon-\sigma^2}+\epsilon}\right). &
\end{aligned}
$$

We denote the event that the above inequality holds to be M. Then we have $\mathbf{P}(M) \geq 1 - \frac{2}{n}$. And hence we have $\mathbf{P}(M^c) \leq \frac{2}{n}$.

Moreover, we also proved that

$$\mathbf{P}(|X| \geq l R_0) \leq n^{-l^2}, \quad l \in \mathbb{Z}_+.$$

Noticing that according to Proposition 15, with probability at least $1-\delta$, we have for any $R \geq 0$ and $t \in [-R, R]$,

$$\left| \tilde{F}_{n,\sigma}^{-1}(F_\sigma(t)) - t \right| \leq 2R + 2\sigma + \frac{KR}{\sigma} + K\sqrt{2\log(2C)} + K\sqrt{2\log\left(\frac{4CRn^2}{\sigma}\right)}.$$

We denote the event that the above inequality holds to be N_δ. Then we have $\mathbf{P}\left(N_\delta\right) \geq$

$1-\delta$. Hence assuming that event N_δ holds, we notice that Proposition 11 indicates

$$
\begin{aligned}
&W_2(\mathbb{P} * \mathcal{N}(0,\sigma^2), \mathbb{P}_n * \mathcal{N}(0,\sigma^2))^2 = \int_{-\infty}^{\infty} \rho_\sigma(t) \left| \tilde{F}_{n,\sigma}^{-1}(F_\sigma(t)) - t \right|^2 dt, \\
&\leq \int_{|t|\leq R_0} \rho_\sigma(t) \left| \tilde{F}_{n,\sigma}^{-1}(F_\sigma(t)) - t \right|^2 dt + \sum_{l=1}^{\infty} \int_{C\cap\{lR_0\leq|t|\leq(l+1)R_0\}} \rho_\sigma(t) \left| \tilde{F}_{n,\sigma}^{-1}(F_\sigma(t)) - t \right|^2 dt \\
&\leq \left(2R_0 + 2\sigma + \frac{KR_0}{\sigma} + K\sqrt{2\log(2C)} + K\sqrt{2\log\left(\frac{4CR_0 n}{\sigma\delta}\right)} \right)^2 \\
&\quad + \sum_{l=2}^{\infty} n^{-l^2} \cdot \left(2lR_0 + 2\sigma + \frac{KlR_0}{\sigma} + K\sqrt{2\log(2C)} + K\sqrt{2\log\left(\frac{4ClR_0 n}{\sigma\delta}\right)} \right)^2 \\
&= \tilde{\mathcal{O}}\left(\log\left(\frac{1}{\delta}\right)\right).
\end{aligned}
$$

Here we use $\tilde{\mathcal{O}}$ to abbreviate the log-items of only n.

Therefore, after concluding all previous upper bounds, we obtain the upper bound on the expectation of $W_2(\mathbb{P} * \mathcal{N}(0,\sigma^2), \mathbb{P}_n * \mathcal{N}(0,\sigma^2))^2$: for any $\epsilon > 0$,

$$
\begin{aligned}
&\mathbb{E}\left[W_2(\mathbb{P} * \mathcal{N}(0,\sigma^2), \mathbb{P}_n * \mathcal{N}(0,\sigma^2))^2\right] \\
&\leq \mathbf{P}(M) \cdot \tilde{\mathcal{O}}\left(n^{-\frac{K^2+\varepsilon}{2K^2+2\varepsilon-\sigma^2}+\epsilon}\right) + \mathbb{E}\left[W_2(\mathbb{P} * \mathcal{N}(0,\sigma^2), \mathbb{P}_n * \mathcal{N}(0,\sigma^2))^2 \mathbf{1}_{M^c}\right] \\
&\leq \tilde{\mathcal{O}}\left(n^{-\frac{K^2+\varepsilon}{2K^2+2\varepsilon-\sigma^2}+\epsilon}\right) + \sum_{j=1}^{\infty} \mathbb{E}\left[W_2(\mathbb{P} * \mathcal{N}(0,\sigma^2), \mathbb{P}_n * \mathcal{N}(0,\sigma^2))^2 \mathbf{1}_{M^c\cap\left(N_{1/n^{j+1}}\setminus N_{1/n^j}\right)}\right] \\
&\quad + \mathbb{E}\left[W_2(\mathbb{P} * \mathcal{N}(0,\sigma^2), \mathbb{P}_n * \mathcal{N}(0,\sigma^2))^2 \mathbf{1}_{M^c\cap N_{1/n}}\right] \\
&\leq \tilde{\mathcal{O}}\left(n^{-\frac{K^2+\varepsilon}{2K^2+2\varepsilon-\sigma^2}+\epsilon}\right) + \sum_{j=1}^{\infty} \mathbf{P}\left(M^c \cap \left(N_{1/n^{j+1}} \setminus N_{1/n^j}\right)\right) \cdot \tilde{\mathcal{O}}\left(\log\left(n^{j+1}\right)\right) \\
&\qquad + \mathbf{P}\left(M^c \cap N_{1/n}\right) \cdot \tilde{\mathcal{O}}\left(\log(n)\right) \\
&\leq \tilde{\mathcal{O}}\left(n^{-\frac{K^2+\varepsilon}{2K^2+2\varepsilon-\sigma^2}+\epsilon}\right) + \sum_{j=1}^{\infty} \left(\frac{1}{n} \wedge \frac{1}{n^j}\right) \cdot \tilde{\mathcal{O}}\left(\log\left(n^{j+1}\right)\right) + \left(\frac{1}{n} \wedge \frac{1}{n^j}\right) \cdot \tilde{\mathcal{O}}\left(\log(n)\right) \\
&= \tilde{\mathcal{O}}\left(n^{-\frac{K^2+\varepsilon}{2K^2+2\varepsilon-\sigma^2}+\epsilon}\right)
\end{aligned}
$$

According to the arbitraity of positive constants ε and ϵ, we can let them both goes to 0, and hence obtain that for any $\epsilon > 0$, we have

$$\mathbb{E}\left[W_2(\mathbb{P} * \mathcal{N}(0, \sigma^2), \mathbb{P}_n * \mathcal{N}(0, \sigma^2))^2\right] = \tilde{\mathcal{O}}\left(n^{-\frac{K^2}{2K^2-\sigma^2}+\epsilon}\right)$$

Therefore, the proof of the upper bound part of Theorem 2 is completed. $\square$

Chapter 6

Proof of Theorem 3

First we start with the definition of Rényi divergence:

Definition 1 (Rényi Divergence and Rényi Mutual Information [12])**.** *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*

$$D_\lambda(P\|Q) \triangleq \frac{1}{\lambda - 1}\log \mathbb{E}_Q\left[\left(\frac{dP}{dQ}\right)^\lambda\right],$$
$$I_\lambda(X;Y) \triangleq D_\lambda(P_{X,Y}\|P_X \otimes P_Y),$$

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.

Lemma 6. *We suppose $(X, Y) \sim P_{X,Y}$, and its marginal distribution to be P_X, P_Y, respectively. We let $\hat{P}_n$ to be an empirical version of P_X generated with n samples. Then for every $1 < \lambda \le 2$, we have*

$$\mathbb{E}[D_{KL}(P_{Y|X} \circ \hat{P}_n\|P_Y)] \le \frac{1}{\lambda - 1}\log(1 + \exp\{(\lambda - 1)(I_\lambda(X;Y) - \log n)\})\,. \tag{6.1}$$

Proof. According to [14], for any distribution P, Q, the function $D_\lambda(P\|Q)$ with respect to $\lambda \in (1, 2]$ is non-decreasing. Hence noticing from [14] that for any distribution

P, Q, $\lim_{\lambda \to 1} D_\lambda(P\|Q) = D_{KL}(P\|Q)$, we have

$$D_{KL}(P_{Y|X} \circ \hat{P}_n \| P_Y) \leq D_\lambda(P_{Y|X} \circ \hat{P}_n \| P_Y).$$

Therefore, it is sufficient to prove that for any $1 < \lambda \leq 2$,

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

We suppose the n samples obtained in 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

$$\begin{aligned}\mathbb{E}[D_\lambda(P_{Y|X} \circ \hat{P}_n \| P_Y)] &= \frac{1}{\lambda - 1} \mathbb{E}\left[\log \mathbb{E}\left[\left\{\frac{d(P_{Y|X} \circ \hat{P}_n)(Y)}{dP_Y(Y)}\right\}^\lambda\right] \middle| X_{1:n}\right] \qquad (6.2)\\ &\leq \frac{1}{\lambda - 1} \log \mathbb{E}\left[\left(\frac{d(P_{Y|X} \circ \hat{P}_n)(Y)}{dP_Y(Y)}\right)^\lambda\right].\end{aligned}$$

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

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

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

$$\mathbb{E}\left[\left\{\frac{dP_{Y|X_{1:n}}(Y|X_{1:n})}{dP_Y(Y)}\right\}^\lambda\right] = \mathbb{E}\left[\left\{\frac{1}{n}\sum_{i=1}^n \frac{dP_{Y|X}(Y|X_i)}{dP_Y(Y)}\right\}^\lambda\right]. \tag{6.3}$$

Moreover, noticing that $(a+b)^{\lambda-1} \le a^{\lambda-1} + b^{\lambda-1}$ holds for $a, b > 0$ and $1 < \lambda \le 2$, we have that for any n *i.i.d.* non-negative random variables B_i $(1 \le i \le n)$,

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

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

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

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

$$\begin{aligned}
&\mathbb{E}\left[\left\{\frac{1}{n}\sum_i \frac{dP_{Y|X}(Y|X_i)}{dP_Y(Y)}\right\}^\lambda \middle| Y\right] \le n^{-\lambda} \cdot \mathbb{E}\left[\left\{\sum_i \frac{dP_{Y|X}(Y|X_i)}{dP_Y(Y)}\right\}^\lambda \middle| Y\right] \\
&\le n^{-\lambda} \cdot \left(n \cdot \mathbb{E}\left[\left\{\frac{dP_{Y|X}(Y|X)}{dP_Y(Y)}\right\}^\lambda \middle| Y\right] + n^\lambda \cdot \left(\mathbb{E}\left[\frac{dP_{Y|X}(Y|X)}{dP_Y(Y)}\middle| Y\right]\right)^\lambda\right) \\
&\le n^{1-\lambda}\mathbb{E}\left[\left\{\frac{dP_{Y|X}(Y|X)}{dP_Y(Y)}\right\}^\lambda \middle| Y\right] + \left(\mathbb{E}\left[\frac{dP_{Y|X}(Y|X)}{dP_Y(Y)}\middle| Y\right]\right)^\lambda.
\end{aligned}$$

Using the fact that $X \perp\!\!\!\perp Y$ and hence $\mathbb{E}[P_{Y|X}(Y|X)|Y] = \int_X P_{Y|X}(Y|X)dP_X(X) = \int_X dP_{X,Y}(X,Y) = P_Y(Y)$, we notice that for any given Y,

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

Therefore, we can upper bound (6.3) as

$$\begin{aligned}
&\mathbb{E}\left[\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{dP_{Y|X}(Y|X_i)}{dP_Y(Y)}\right\}^\lambda\right] = \mathbb{E}\left[\mathbb{E}\left[\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{dP_{Y|X}(Y|X_i)}{dP_Y(Y)}\right\}^\lambda\right]\middle|Y\right] \\
&\leq n^{1-\lambda}\mathbb{E}\left[\mathbb{E}\left[\left\{\frac{dP_{Y|X}(Y|X)}{dP_Y(Y)}\right\}^\lambda\middle|Y\right]\middle|Y\right] + \mathbb{E}\left[\left(\mathbb{E}\left[\frac{dP_{Y|X}(Y|X)}{dP_Y(Y)}\middle|Y\right]\right)^\lambda\middle|Y\right] \\
&\leq n^{1-\lambda}\mathbb{E}\left[\left\{\frac{dP_{Y|X}(Y|X)}{dP_Y(Y)}\right\}^\lambda\right] + 1 \\
&= n^{1-\lambda}\cdot\exp\left((\lambda-1)I_\lambda(X;Y)\right) + 1.
\end{aligned}$$

This implies that

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

which together with (6.2) recovers (6.1). □

Remark 8. *Hayashi [9] upper bounds the LHS of* (6.1) *with*

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

where $K_\lambda(X;Y) = \inf_{Q_Y} D_\lambda(P_{X,Y}\|P_X Q_Y)$ *is the so-called Sibson-Csiszar information, cf. [13]. 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 [7]. We note that [9, 8] also contain bounds on* $\mathbb{E}[\mathrm{TV}(P_{Y|X} \circ \hat{P}_n, P_Y)]$ *which do not assume existence of* $\lambda > 1$ *moment of* $\frac{P_{Y|X}}{P_Y}$ *and instead rely on the distribution of* $\log\frac{dP_{Y|X}}{dP_Y}$.

Lemma 7. *Suppose* $\mathbb{P}$ *is a* d*-dimensional* K*-subgaussian distribution and random variables* $X \sim \mathbb{P}, Z \sim \mathcal{N}(0, \sigma^2 I_d)$ *are independent to each other. We let* $Y = X + Z$.

Then for any $\sigma > 0$ and $1 < \lambda < 2$, there exists a positive constant C only depending on $\mathbb{P}$ and K, σ such that

$$I_\lambda(X;Y) \le \frac{1}{\lambda - 1} \log\left(\frac{C}{(2-\lambda)^{d+1}}\right).$$

Proof. We use P_X, P_Y and $P_{X,Y}, P_{Y|X}$ to denote the marginal distributions with respect to X, Y, the joint distribution of (X, Y) and conditional distribution of Y given X. According to the definition of Rényi mutual information, we have

$$I_\lambda(X;Y) = \frac{1}{\lambda - 1} \log\left(\mathbb{E}_{P_X \otimes P_Y}\left[\left(\frac{dP_{X,Y}}{d(P_X \otimes P_Y)}\right)^\lambda\right]\right).$$

Denoting the PDFs of distributions $P_Y, P_{Y|X}$ as $\rho_Y(\cdot), \rho_{Y|X}(\cdot)$, we have

$$\rho_{Y|X}(y|X) = \varphi_{\sigma^2 I_d}(y - X), \quad \rho_Y(y) = \mathbb{E}[\varphi_{\sigma^2 I_d}(y - \tilde{X})].$$

If we choose $\tilde{X} \sim \mathbb{P}, \tilde{X} \perp\!\!\!\perp X$, then we have

$$\begin{aligned}
\mathbb{E}_{P_X \otimes P_Y}\left[\left(\frac{dP_{X,Y}}{d(P_X \otimes P_Y)}\right)^\lambda\right] =& \mathbb{E}\left[\left(\frac{dP_{X,Y}(X,Y)}{dP_X \otimes P_Y(X,Y)}\right)^\lambda\right] \\
=& \mathbb{E}\left[\mathbb{E}\left[\left(\frac{dP_{Y|X}(Y|X)}{dP_Y(Y)}\right)^\lambda \middle| X\right]\right] \\
=& \mathbb{E}\left[\mathbb{E}\left[\left(\frac{\rho_{Y|X}(Y|X)}{\rho_Y(Y)}\right)^\lambda \middle| X\right]\right] \\
=& \mathbb{E}\left[\int_{\mathbb{R}^d} \rho_Y(y) \left(\frac{\rho_{Y|X}(y|X)}{\rho_Y(y)}\right)^\lambda dy\right] \\
=& \mathbb{E}\left[\int_{\mathbb{R}^d} \frac{\varphi_{\sigma^2 I_d}(y - X)^\lambda}{(\mathbb{E}[\varphi_{\sigma^2 I_d}(y - \tilde{X})])^{\lambda-1}} dy\right].
\end{aligned}$$

Moreover, Noticing that

$$\varphi_{\sigma^2 I_d}(x) = \frac{1}{(\sqrt{2\pi}\sigma)^d} \exp\left(-\frac{\|x\|_2^2}{2\sigma^2}\right), \quad \forall x \in \mathbb{R}^d,$$

we have

$$\mathbb{E}\left[\int_{\mathbb{R}^d} \frac{\varphi_{\sigma^2 I_d}(y-X)^\lambda}{(\mathbb{E}[\varphi_{\sigma^2 I_d}(y-\tilde{X})])^{\lambda-1}} dy\right] = \frac{1}{(\sqrt{2\pi}\sigma)^d} \cdot \mathbb{E}\left[\int_{\mathbb{R}^d} \frac{\exp\left(-\lambda\|y-X\|_2^2/2\right)}{(\mathbb{E}\exp(-\|y-\tilde{X}\|_2^2/2))^{\lambda-1}} dy\right]$$

Therefore, we only need to prove that there exist positive constant $C = C(\mathbb{P}, K, \sigma)$ such that

$$\mathbb{E}\left[\int_{\mathbb{R}^d} \frac{\exp\left(-\lambda\|y-X\|_2^2/2\right)}{(\mathbb{E}\exp(-\|y-\tilde{X}\|_2^2/2))^{\lambda-1}} dy\right] \leq \frac{C}{(2-\lambda)^{d+1}} \cdot (\sqrt{2\pi}\sigma)^d \tag{6.4}$$

WLOG, we assume $\sigma = 1$ (otherwise we substitute K with K/σ and let $\sigma = 1$). We let $A_k = \{X | \|X\|_2 \in [k, k+1)\} \subset \mathbb{R}^d$. Then we have

$$\mathbb{R}^d = \bigcup_{k=0}^{\infty} A_k,$$

Let m_k to be smallest number of l_2-balls with diameter 2 in $\mathbb{R}^d$ which can cover the set A_k. Then we have

$$m_k = C_1(k+1)^d$$

for some positive constant C_1 (note that here we only need to prove that the integral is finite, hence we ignore the constants). We use $A_{k,1}, \cdots, A_{k,m_k}$ to denote the intersection between each of these m_k balls with A_k. Then we have

$$A_k = \bigcup_{i=1}^{m_k} A_{k,i}, \quad \text{and} \quad \operatorname{diam}(A_{k,i}) \leq 2, \quad \forall 1 \leq i \leq m_k.$$

Assuming $X \sim \mathbb{P}$, we denote

$$p_{k,i} = \mathbf{P}(X \in A_{k,i}), \quad p_k = \mathbf{P}(X \in A_k).$$

Since for any $X, \tilde{X} \in A_{k,i}$ we have $\|X - \tilde{X}\|_2 \le 2$, we obtain that $\forall k, i$ and $X \in A_{k,i}$,

$$\begin{aligned}
&\mathbb{E}\exp\left(-\frac{\|y-\tilde{X}\|^2}{2}\right) \ge \mathbb{E}\left[\mathbf{1}_{\tilde{X}\in A_{k,i}}\exp\left(-\frac{\|y-\tilde{X}\|^2}{2}\right)\right] \\
&\ge p_{k,i}\min_{\tilde{X}\in A_{k,i}}\exp\left(-\frac{\|y-\tilde{X}\|^2}{2}\right) = p_{k,i}\min_{\tilde{X}\in A_{k,i}}\exp\left(-\frac{\|y-X+(X-\tilde{X})\|^2}{2}\right) \\
&\ge p_{k,i}\exp\left(-\frac{(\|y-X\|+2)^2}{2}\right).
\end{aligned}$$

Noticing the fact that $\lambda < 2$, we have

$$\begin{aligned}
&\mathbb{E}\left[\mathbf{1}_{X\in A_{k,i}}\int_{\mathbb{R}^d}\frac{\exp(-\lambda\|y-X\|^2/2)}{(\mathbb{E}\exp(-\|y-\tilde{X}\|^2/2))^{\lambda-1}}dy\right] \\
&\le \mathbb{E}\left[\mathbf{1}_{X\in A_{k,i}}\int_{\mathbb{R}^d}\frac{\exp(-\lambda\|y-X\|^2/2)}{p_{k,i}^{\lambda-1}\exp\left(-(\|y-X\|+2)^2/2\right)^{\lambda-1}}dy\right] \\
&\le \mathbb{E}\left[\frac{\mathbf{1}_{X\in A_{k,i}}}{p_{k,i}^{\lambda-1}}\int_{\mathbb{R}^d}\exp\left(-\frac{\lambda\|y-X\|^2}{2}+\frac{(\lambda-1)(\|y-X\|+2)^2}{2}\right)dy\right] \\
&= \mathbb{E}\left[\frac{\mathbf{1}_{X\in A_{k,i}}}{p_{k,i}^{\lambda-1}}\int_{\mathbb{R}^d}\exp\left(-\frac{1}{2}\|u\|^2+2(\lambda-1)\|u\|+2(\lambda-1)\right)du\right] \\
&\le \mathbb{E}\left[\frac{\mathbf{1}_{X\in A_{k,i}}}{p_{k,i}^{\lambda-1}}\int_{\mathbb{R}^d}\exp\left(-\frac{1}{2}\|u\|^2+2\|u\|+2\right)du\right] \\
&= p_{k,i}^{2-\lambda}\cdot\int_{\mathbb{R}^d}\exp\left(-\frac{1}{2}\|u\|^2+2\|u\|+2\right)du.
\end{aligned}$$

Let constant $C_2 = \int_{\mathbb{R}^d}\exp\left(-\frac{1}{2}\|u\|^2+2\|u\|+2\right)du < \infty$, and noticing that $A_{k,i} \subset A_k$ for each i, we obtain that

$$\mathbb{E}\left[\mathbf{1}_{X\in A_{k,i}}\int_{\mathbb{R}^d}\frac{\exp(-\lambda\|y-X\|^2/2)}{(\mathbb{E}\exp(-\|y-\tilde{X}\|^2/2))^{\lambda-1}}dy\right] \le C_2 p_{k,i}^{2-\lambda} \le C_2 p_k^{2-\lambda}.$$

Hence,

$$\mathbb{E}\left[\mathbf{1}_{X\in A_k}\int_{\mathbb{R}^d}\frac{\exp(-\lambda\|y-X\|^2/2)}{(\mathbb{E}\exp(-\|y-\tilde{X}\|^2/2))^{\lambda-1}}dy \le C_2 m_k p_k^{2-\lambda}\right] \le C_1 C_2 (k+1)^d p_k^{2-\lambda}.$$

Since $\mathbb{P}$ is a K-subgaussian random variable, we have

$$p_k = \mathbf{P}(\|X\|_2 \in [k, k+1)) \leq \mathbf{P}(\|X\|_2 \geq k) \leq C_0 \exp\left(-\frac{k^2}{2K^2}\right).$$

Therefore, we obtain that

$$\mathbb{E}\left[\mathbf{1}_{X\in A_k} \int_{\mathbb{R}^d} \frac{\exp(-\lambda\|y-X\|^2/2)}{(\mathbb{E}\exp(-\|y-\tilde{X}\|^2/2))^{\lambda-1}} dy\right] \leq C_0C_1C_2(k+1)^d \exp\left(-\frac{(2-\lambda)k^2}{2K^2}\right),$$

and hence

$$\begin{aligned}
\mathbb{E}\left[\int_{\mathbb{R}^d} \frac{\exp(-\lambda\|y-X\|^2/2)}{(\mathbb{E}\exp(-\|y-\tilde{X}\|^2/2))^{\lambda-1}} dy\right] &\leq \sum_{k=0}^{\infty} C_0C_1C_2(k+1)^d \exp\left(-\frac{(2-\lambda)k^2}{2K^2}\right) \\
&\leq \sum_{k=0}^{\infty} C_0C_1C_2(k+1)^d \exp\left(-\frac{(2-\lambda)k}{2K^2}\right) \\
&\leq C_0C_1C_2 d! \cdot \left(1 - \exp\left(-\frac{2-\lambda}{2K^2}\right)\right)^{-d-1}.
\end{aligned}$$

Here in the second inequality we use the fact that $\sum_{k=0}^{\infty}(k+1)^d c^{-k} \leq d! c^{-k-1}$ for any $0 < c < 1$. Next noticing that $1 - \exp(-x) \leq 1 - (1-x) = x$ holds for all $x \in \mathbb{R}$, we obtain that

$$\mathbb{E}\left[\int_{\mathbb{R}^d} \frac{\exp(-\lambda\|y-X\|^2/2)}{(\mathbb{E}\exp(-\|y-\tilde{X}\|^2/2))^{\lambda-1}} dy\right] \leq C_0C_1C_2 d! \left(\frac{2K^2}{2-\lambda}\right)^{d+1} \leq \frac{C}{(2-\lambda)^{d+1}} \cdot (\sqrt{2\pi}\sigma)^d$$

with $C = C_0C_1C_2 d!(2K^2)^{d+1}/(\sqrt{2\pi}\sigma)^d$. Hence (6.4) is verified. $\square$

Equipped with the above lemmas, we are 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 we have $P_{Y|X} \sim \mathcal{N}(X, \sigma^2 I_d)$, which indicates that $P_{Y|X} \circ \mathbb{P}_n \sim \mathbb{P}_n * \mathcal{N}(0, \sigma^2 I_d)$.

Therefore, adopting Lemma 6 and Lemma 7, we obtain that for any $1 < \lambda < 2$,

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

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

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

$\square$

Chapter 7

Conclusion

As the convergence from the empirical measure $\hat{\mathbb{P}}_n$ to the population measure $\mathbb{P}$ under Wasserstein distance always suffers from the curse of dimensionality. People seek to resolve this problem using the convergence from the smoothed empirical measure $\hat{\mathbb{P}}_n * \mathcal{N}(0, \sigma^2 I_d)$ to the smoothed population measure $\mathbb{P}_n * \mathcal{N}(0, \sigma^2 I_d)$. However, the exact convergence rate of the smoothed empirical measure is not perfectly understood till this paper.

Suppose $\mathbb{P}$ is a K-subGaussian distribution, we prove a dichotomy of the convergence rate under W2 distance squared when $K < \sigma$ and when $K > \sigma$, i.e. when $K < \sigma$ the convergence rate is at $\mathcal{O}(1/n)$ and when $K > \sigma$ there exists a case such that the convergence rate is of $\omega(1/n)$. Moreover, for 1D cases, we provide detailed analysis on the convergence rate when $K > \sigma$, which is always the case when the convolution with Gaussian been viewed as adding a noise of small scale. Specifically, we prove that the convergence rate changes gradually from $1/\sqrt{n}$ to $1/n$ as σ/K goes from zero to one.

Beyond W2 distance, we also proved that the convergence rate under KL divergence is always $\mathcal{O}(1/n)$, as long as $\sigma > 0$. This indicates that the convergence rate of KL divergence is faster than the convergence rate of W2 distance when $\sigma < K$, indicating a failure of T2 inequality for $\mathbb{P} * \mathcal{N}(0, \sigma^2)$ when $K > \sigma$.

Appendix A

Proof of Subgaussianity in Section 3

Proposition 16. *Given positive constant $c > 2, c_1 > 0$, we consider 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*

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

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 $S, \tilde{S} \sim \mathbb{P}, S \perp\!\!\!\perp \tilde{S}$,

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

Proof. We let

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

(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

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

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 $\alpha K^2 \leq -2S_1$, we have

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

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

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

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

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

Hence if

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

we would have

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

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

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

and

$$\begin{aligned}
h'(\alpha) &= -p_0 K^2 + c_1 \sum_{k=1}^{\infty} \left(r_k^2 - \alpha K^2 r_k - S_1 r_k - K^2\right) \exp\left(-\frac{r_k^2}{2K^2} + \alpha r_k\right) \\
&\leq -p_0 K^2 + c_1 \sum_{k=1}^{\infty} \left(r_k^2 - \alpha K^2 r_k\right) \exp\left(-\frac{r_k^2}{2K^2} + \alpha r_k\right) \\
&\leq -p_0 K^2 + c_1 \sum_{k=1}^{\infty} \left(r_k^2 - \alpha K^2 r_k\right) \exp\left(-\frac{r_k^2}{2K^2} + \frac{r_k}{4K^2}\right) \\
&\leq -p_0 K^2 + 2c_1 \sum_{k=1}^{\infty} r_k^2 \exp\left(-\frac{r_k^2}{4K^2}\right),
\end{aligned}$$

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

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

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

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

Therefore, we have

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

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

$$h(0) = p_0 S_1 + c_1 \sum_{k=1}^{\infty} (r_k - S_1) \exp\left(-\frac{r_k^2}{2K^2}\right) = p_0 S_1 + \sum_{k=1}^{\infty} p_k (r_k - S_1) = \mathbb{E}[S] - S_1 = 0.$$

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

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

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

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

1. $2S_1 \leq 1$, which can be obtained from $c_1 \leq \frac{\left(1-\exp\left(-\frac{1}{2K^2}\right)\right)^2}{2\exp\left(-\frac{1}{2K^2}\right)}$;
2. $c_1 \leq \frac{1}{24}$;
3. $c_1 \leq \frac{1}{2}\left(1 - \exp\left(-\frac{1}{8K^2}\right)\right)\left(1 - \exp\left(-\frac{1}{2K^2}\right)\right)$;
4. $1 - p_0 = \sum_{k=1}^{\infty} p_k \leq \frac{1}{2}$, which can be obtained from $c_1 \leq \frac{1-\exp\left(-\frac{1}{2K^2}\right)}{2}$.

Hence if we choose

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

and p_k in (3.1), we would have $f(\alpha) \leq 1$ for all $\alpha \in \mathbb{R}$. Therefore, we have

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

which indicates that distribution P is a K-subgaussian. $\square$

Appendix B

LSI and T_2 constants for Bernoulli-Gaussian mixtures

B.0.1 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

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

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

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

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

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

which indicates that

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

We use $\varphi_\sigma(x)$ 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

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

and also

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

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

$$C_h \geq \lim_{x_1 \to -\infty} \frac{q_{h,3} q_{h,5}}{q_{h,2} + q_{h,4}} - 1 = \lim_{x_1 \to -\infty} \frac{q_3 q_5}{q_4} - 1 = \frac{(1 - q_4 - q_5) q_5}{q_4} - 1 \geq \frac{(1 - q_5) q_5}{q_4} - 2.$$

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

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

and also

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

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.

□

B.0.2 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

$$W_2(\mathbb{P}_h * \mathcal{N}(0,\sigma^2), \mathbb{Q}) \leq C'_h D_{KL}(\mathbb{P}_h * \mathcal{N}(0,\sigma^2) \| \mathbb{Q}) \qquad \forall \textit{ distribution } \mathbb{Q}. \tag{B.1}$$

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 $(1-\delta)(1+\sigma^2/K^2)^2h^2 > 4\sigma^2/K^2$, and $\mathbb{Q}^\sigma_h = \mathbb{Q}_h * \mathcal{N}(0,\sigma^2)$.

According to data-processing inequality we have

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

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 $\tilde{F}_q(t) - \tilde{F}_p(t) = (q-p)(\Phi_\sigma(t) - \Phi_\sigma(t-h))$ where $\hat{F}_q, \hat{F}_p, \Phi_\sigma$ are CDFs of distribution $\mathbb{Q} * \mathcal{N}(0,\sigma^2), \mathbb{P} * \mathcal{N}(0,\sigma^2), \mathcal{N}(0,\sigma^2)$. We can prove that

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

while

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

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

Bibliography

[1] 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.

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

[3] 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.

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

[5] 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.

[6] 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.

[7] 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.

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

[9] 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.

[10] Soheil Kolouri, Se Rim Park, Matthew Thorpe, Dejan Slepcev, and Gustavo K Rohde. Optimal mass transport: Signal processing and machine-learning applications. *IEEE signal processing magazine*, 34(4):43–59, 2017.

[11] Sloan Nietert, Ziv Goldfeld, and Kengo Kato. From smooth wasserstein distance to dual sobolev norm: Empirical approximation and statistical applications. *arXiv preprint arXiv:2101.04039*, 2021.

[12] 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.

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

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

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

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

[17] 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.

[18] 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.

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