Tail bound for sup-norm distance

Jiaxin Hu

April 3, 2022

We first show the key Lemma for the proof of unseeded algorithm, which indicates the tail bound for the sup-norm distance between empirical distributions. The result is worse than the first inequality in Conjecture 1 in 0306.

Notation need to be added:

- Let $|I|$ denotes the cardinality of the index set I.

Lemma 1 (Tail bound for the sup-norm distance)**.** *Suppose that we have i.i.d. samples* $(X_1, Y_1), \ldots,$ (X_n, Y_n) *following the multivariate zero-mean Gaussian distribution with variance 1 and correlation* $\rho \in [0, 1)$*; i.e,*

$$(X_i, Y_i) \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \quad \textit{and} \quad (X_i, Y_i) \perp (X_j, Y_j), \textit{ for all } i \neq j. \tag{1}$$

Define the sup-norm distance for the empirical distributions as

$$d = \sup_{t \in \mathbb{R}} |F_n(t) - G_n(t)|,$$

where

$$F_n(t) = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}\{X_i \leq t\}, \quad G_n(t) = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}\{Y_i \leq t\}$$

are empirical distributions for X and Y, respectively. Then, for all $a > 0$, there exists a positive constant C such that

$$\mathbb{P}(d > a) \leq C n^2 \exp\left(-\frac{n a^2}{\sqrt{1 - \rho^2}}\right).$$

Remark 1. Let $\sigma^2 = \sqrt{1 - \rho^2}$ and $a = \sqrt{\sigma / n}$. We have

$$\mathbb{P}\left(d > \sqrt{\frac{\sigma}{n}}\right) \leq C n^2 \exp(-1).$$

The right hand side is larger than $\mathcal{O}(\exp(-\sigma^{-1}))$ that we expect in Conjecture 1, and we can not guarantee the performance of unseeded algorithm under any condition of σ.

Instead, take $a = \sqrt{\sigma^{1/2}/n}$. By current Lemma 1 and the small ball probability (14) in Ding et al. (2021), for the true pair (i, k) and fake pair (i', k'), we have

$$\mathbb{P}\left(d_{ik} > \sqrt{\frac{\sigma^{1/2}}{n}}\right) \leq C_1 n^2 \exp(-\sigma^{-1/2}), \quad \text{and} \quad \mathbb{P}\left(d_{i'k'} \leq \sqrt{\frac{\sigma^{1/2}}{n}}\right) \leq C_2 \exp(-\sigma^{-1/2}).$$

Then, we are able to guarantee the unseeded algorithm when $\sigma \lesssim \log^{-2} n$, which is a stricter condition than Ding's result $\sigma \lesssim \log^{-1} n$.

Proof of Lemma 1. By Chernoff bound, for all $\lambda > 0$ we have

$$\mathbb{P}(d > a) \leq \exp(-\lambda a)\mathbb{E}[\exp(\lambda d)]. \tag{2}$$

Our proof idea is to bound the expectation $\mathbb{E}[\exp(\lambda d)]$ and find an optimal λ tightening the tail probability. We will use symmetrization technique and conditional expectation to show that controlling $\mathbb{E}[\exp d]$ is equivalent to controlling the maximal expectation $\sup_{t\in\mathbb{R}} \mathbb{E}[\exp|\{i : X_i \leq t, Y_i > t\}|]$ which is a function of the correlation ρ.

Now, we investigate the expectation $\mathbb{E}[\exp(\lambda d)]$ in detail.

First, we adapt symmetrization technique by considering the variables $Z_i \coloneqq (X_i, Y_i)$ for all $i \in [n]$, and the family $\mathcal{F} \coloneqq \{f_t : \mathbb{R}^2 \mapsto \mathbb{R}\colon f_t(x, y) = \frac{1}{n}(\mathbb{1}\{x \leq t\} - \mathbb{1}\{y \leq t\}), t \in \mathbb{R}\}$. Note that for all $t \in \mathbb{R}$, we have $F_n(t) - G_n(t) = \sum_{i\in[n]} f_t(Z_i)$, $\mathbb{E}[f_t(Z_i)] = 0$, and $\mathbb{E}[|f_t(Z_i)|] \leq 2$. Then, we have

$$\begin{aligned}
\mathbb{E}[\exp(\lambda d)] &\leq 2\mathbb{E}\left[\sup_{t\in\mathbb{R}} \exp\left(\lambda[F_n(t) - G_n(t)]\right)\right] \\
&= 2\mathbb{E}\left[\sup_{f_t\in\mathcal{F}} \exp\left(\lambda \sum_{i\in[n]} (f_t(Z_i) - \mathbb{E}[f_t(Z_i)])\right)\right] \\
&\leq 2\mathbb{E}\left[\sup_{f_t\in\mathcal{F}} \exp\left(2\lambda \sum_{i\in[n]} f_t(Z_i)\epsilon_i\right)\right],
\end{aligned} \tag{3}$$

where the first inequality follows from the fact that $\mathbb{E}[|X|] \leq 2\mathbb{E}[X]$ holds for any random variable X with symmetric density around 0, the last inequality follows Symmetrization Lemma 2 taking the measurable space $(\mathbb{X}, \mathcal{X}) = (\mathbb{R}^2, \mathcal{B}^2)$ and convex function $\Psi(x) = \exp(\lambda x)$, and ϵ_i's are i.i.d. Rademacer variables which take values $\{-1, +1\}$ equiprobably and independent with Z_i's .

Second, we bound the right hand side of (3) by controlling the conditional expectation given Z_i's. Before the derivation of upper bound, we consider following definitions and facts.

(a) [**Decomposition.**] Define the index sets

$$L_Z(t) = \{i : X_i > t, Y_i \leq t, i \in [n]\} \quad \text{and} \quad R_Z(t) = \{i : X_i \leq t, Y_i > t, i \in [n]\},$$

where $L_Z(t)$ and $R_Z(t)$ collect the indices of Z_i when $f_t(Z_i)$ is -1 and 1, respectively. The summation in (3) satisfies

$$\sum_{i\in[n]} f_t(Z_i)\epsilon_i = -\frac{1}{n} \sum_{i\in L_Z(t)} \epsilon_i + \frac{1}{n} \sum_{i\in R_Z(t)} \epsilon_i.$$

(b) [**Independence and identical distribution of** $R_Z(t), L_Z(t)$**.**] The sets $L_Z(t) \cap R_Z(t) = \emptyset$ and thus $\{\epsilon_i\}_{i \in L_Z(t)}$ are independent with $\{\epsilon_i\}_{i \in R_Z(t)}$. The distributions of $L_Z(t)$ and $R_Z(t)$ with respect to Z are identical for all $t \in \mathbb{R}$.

(c) [**Collections of** $R_Z(t), L_Z(t)$**.**] Let $R_Z = \{R_Z(t) : t \in \mathbb{R}\}$ and $L_Z = \{L_Z(t) : t \in \mathbb{R}\}$ be the collections of all $R_Z(t)$ and $L_Z(t)$ with given $\{Z_i\}_{i \in [n]}$. For any $\{Z_i\}_{i \in [n]}$, the cardinalities of the collections $|R_Z|, |L_Z| \leq 2n$ because the set $\{X_1, \ldots, X_n, Y_1, \ldots, Y_n\}$ has at most $2n$ different values.

(d) [**Maxima and summation.**] For any positive-valued random variables X_i and Y_i such that X_i and Y_i are independent for all $i \in [n]$, following inequalities hold

$$\mathbb{E}[\max_{i \in [n]} X_i Y_i] \leq \mathbb{E}[\max_{i \in [n]} X_i] \mathbb{E}[\max_{i \in [n]} Y_i] \quad \text{and} \quad \mathbb{E}[\max_{i \in [n]} X_i] \leq \sum_{i \in [n]} \mathbb{E}[X_i]. \tag{4}$$

The inequalities also hold by replacing $\mathbb{E}[\cdot]$ to $\mathbb{E}[\cdot|Z_i]$, where Z_i is independent with X_i, Y_i.

(e) [**Rademacher variables.**] Let ϵ_i's are i.i.d. Rademacer variables which take values $\{-1, +1\}$ equiprobably where $i \in [n]$. Then, we have

$$\mathbb{E}[\exp(\theta \sum_{i \in [n]} \epsilon_i)] \leq \exp\left(\frac{\theta^2 n}{2}\right), \quad \text{for all} \quad \theta > 0.$$

Then, let Z be the shorthand of $\{Z_i\}_{i \in [n]}$. We have the conditional expectation

$$\begin{aligned}
&\mathbb{E}\left[\sup_{f_t \in \mathcal{F}} \exp\left(2\lambda \sum_{i \in [n]} f_t(Z_i)\epsilon_i\right) |Z\right] \\
&= \mathbb{E}\left[\sup_{t \in \mathbb{R}} \exp\left(-\frac{2\lambda}{n} \sum_{i \in L_Z(t)} \epsilon_i\right) \exp\left(\frac{2\lambda}{n} \sum_{i \in R_Z(t)} \epsilon_i\right) |Z\right] \\
&\leq \mathbb{E}\left[\max_{L_Z(t) \in L_Z} \exp\left(-\frac{2\lambda}{n} \sum_{i \in L_Z(t)} \epsilon_i\right) |Z\right] \mathbb{E}\left[\max_{R_Z(t) \in R_Z} \exp\left(\frac{2\lambda}{n} \sum_{i \in R_Z(t)} \epsilon_i\right) |Z\right] \\
&\leq \left(\sum_{L_Z(t) \in L_Z} \mathbb{E}\left[\exp\left(-\frac{2\lambda}{n} \sum_{i \in L_Z(t)} \epsilon_i\right) |Z\right]\right) \cdot \left(\sum_{R_Z(t) \in R_Z} \mathbb{E}\left[\exp\left(\frac{2\lambda}{n} \sum_{i \in R_Z(t)} \epsilon_i\right) |Z\right]\right) \\
&\leq \left(\sum_{L_Z(t) \in L_Z} \exp\left(\frac{2\lambda^2}{n^2} |L_Z(t)|\right)\right) \cdot \left(\sum_{R_Z(t) \in R_Z} \exp\left(\frac{2\lambda^2}{n^2} |R_Z(t)|\right)\right),
\end{aligned} \tag{5}$$

where the equation follows from observation (a), the first inequality follows from (b) and the first inequality in (4) of (d), the second inequality follows from the second inequality in (4) of (d), and the last inequality follows from observation (e).

Next, we plug the upper bound for conditional expectation (5) back to the upper bound for the expectation (3). We have

$$\mathbb{E}[\exp(\lambda d)] \leq 2\mathbb{E}\left[\mathbb{E}\left[\sup_{f_t \in \mathcal{F}} \exp\left(2\lambda \sum_{i \in [n]} f_t(Z_i)\epsilon_i\right) |Z\right]\right]$$

$$
\begin{aligned}
&\leq 2\mathbb{E}\left[\left(\sum_{L_Z(t)\in L_Z} \exp\left(\frac{2\lambda^2}{n^2}|L_Z(t)|\right)\right)\cdot\left(\sum_{R_Z(t)\in R_Z} \exp\left(\frac{2\lambda^2}{n^2}|R_Z(t)|\right)\right)\right] \\
&\leq 2\left(\mathbb{E}\left[\sum_{R_Z(t)\in R_Z} \exp\left(\frac{2\lambda^2}{n^2}|R_Z(t)|\right)\right]\right)^2 \\
&\leq 2\left(2n\sup_{t\in\mathbb{R}}\mathbb{E}\left[\exp\left(\frac{2\lambda^2}{n^2}|R_Z(t)|\right)\right]\right)^2 \\
&\leq 8n^2\exp\left(\frac{4\lambda^2}{n}\sqrt{1-\rho^2}\right),
\end{aligned}
\tag{6}
$$

where the second inequality follows from observation (b), the third inequality follows from the observation (c) and the fact that $\mathbb{E}[\exp(\theta|R_Z(t)|)] \leq \sup_{t\in\mathbb{R}}\mathbb{E}[\exp(\theta|R_Z(t)|)]$ for all $t\in\mathbb{R}$ and $\theta>0$, and the last inequality follows from Proposition 1.

Last, we plug the upper bound (6) back to the Chernoff bound (2). Taking $\lambda = na/(8\sqrt{1-\rho^2})$, we obtain that

$$
\mathbb{P}(d>a) \leq 8n^2\exp\left(\frac{4\lambda^2}{n}\sqrt{1-\rho^2}-\lambda a\right) = 8n^2\exp\left(-\frac{na^2}{16\sqrt{1-\rho^2}}\right).
$$

□

Lemma 2 (Symmetrization Lemma)**.** *Assume that $\Psi:\mathbb{R}\mapsto\mathbb{R}_+$ is a convex function, $Z_1,\ldots,Z_n$ are i.i.d. random variables with values in the measurable space $(\mathbb{X},\mathcal{X})$, and $\mathcal{F}$ is a countable family of measurable function from $\mathbb{X}$ to $\mathbb{R}$ with $\mathbb{E}[|f(Z_1)|]<\infty$ for all $f\in\mathcal{F}$. Then, we have*

$$
\mathbb{E}\left[\sup_{f\in\mathcal{F}}\Psi\left(\sum_{i\in[n]}(f(Z_i)-\mathbb{E}[f(Z_i)])\right)\right] \leq \mathbb{E}\left[\sup_{f\in\mathcal{F}}\Psi\left(2\sum_{i\in[n]}f(Z_i)\epsilon_i\right)\right],
$$

where $\epsilon_1,\ldots,\epsilon_n$ are i.i.d. Rademacer variables which take values $\{-1,+1\}$ equiprobably, and $\epsilon_i\perp Z_j$ for all $i,j\in[n]$.

Proof of Lemma 2. See the prove of Lemma 2 in the reference. □

Proposition 1. *Suppose that we have samples $(X_1,Y_1),\ldots,(X_n,Y_n)$ from (1); i.e., (X_i,Y_i) i.i.d. follow the multivariate zero-mean Gaussian distribution with variance 1 and correlation $\rho\in(0,1)$. For all $t\in\mathbb{R}$, define the set*

$$
R(t)=\{i: X_i\leq t, Y_i>t, i\in[n]\}.
$$

Then, the cardinality $|R(t)|\sim Bin(n,p(t))$, where

$$
p(t)=\mathbb{P}(X_1\leq t, Y_1>t), \quad \textit{and} \quad p(t)\leq\sqrt{1-\rho^2}.
$$

Further, we have the moment generating function

$$
\mathbb{E}[\exp(\theta|R(t)|)] \leq \exp(n\theta\sqrt{1-\rho^2}), \quad \textit{for all} \quad \theta>0.
$$

Proof of Proposition 1. By the definition, for all $t \in \mathbb{R}$, we have

$$|R(t)| = \sum_{i \in [n]} \mathbb{1}\{X_i \leq t, Y_i > t\},$$

where the indicators $\mathbb{1}\{X_i \leq t, Y_i > t\}$ follows the Bernoulli distribution with parameter $p(t) = \mathbb{P}(X_i \leq t, Y_i > t)$, independently, and thus $|R(t)| \sim \mathrm{Bin}(n, p(t))$.

Now, we find the upper bound for $p(t)$. Note that $Y_1 = \rho X_1 + \sqrt{1-\rho^2} W$ where W is a standard Gaussian variable independent with X_1. Then, we have

$$\begin{aligned} p(t) &= \mathbb{P}\left(X_1 \leq t, W > \frac{t - \rho X_1}{\sqrt{1-\rho^2}}\right) \\ &= \int_{-\infty}^{t} \mathbb{P}\left(W > \frac{t - \rho x}{\sqrt{1-\rho^2}}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\ &\leq \sqrt{1-\rho^2} \exp(-t^2) \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{(x-\rho t)^2}{2(1-\rho^2)}\right) dx \qquad (7) \\ &\leq \sqrt{1-\rho^2}, \end{aligned}$$

where the first inequality follows from the fact that $\mathbb{P}(W \geq t) \leq \exp(-t^2/2)$ for any standard Gaussian variable W, and the last inequality follows by observing that the integral in (7) is equivalent to the probability $\mathbb{P}(X < t)$ for $X \sim N(\rho t, 1-\rho^2)$ and $\exp(-t^2) \leq 1$ for all $t \in \mathbb{R}$.

Last, we have

$$\mathbb{E}[\exp(\theta |R(t)|)] = \left[1 + \frac{np(t)(e^\theta - 1)}{n}\right]^n \leq \exp\left(n\theta p(t)\right) \leq \exp\left(n\theta\sqrt{1-\rho^2}\right),$$

where the equation follows from the moment generating function for the Binomial variable, the first inequality follows from the basic inequalities $(1 + x/n)^n \leq e^x$ and $e^x \geq 1 + x$, and the last inequality follows from previous result $p(t) \leq \sqrt{1-\rho^2}$. $\square$

References

Ding, J., Ma, Z., Wu, Y., and Xu, J. (2021). Efficient random graph matching via degree profiles. *Probability Theory and Related Fields*, 179(1):29–115.