Conjecture 1

Jiaxin Hu

March 19, 2022

Conjecture 1 (Tail bounds for empirical process)**.** Consider the correlated pairs of normal variables (X_i, Y_i) for $i \in [n]$, where $X_i, Y_i \sim N(0,1)$ and $\text{cov}(X_i, Y_i) = \rho$. Let $\rho = \sqrt{1-\sigma^2}$, and F_n, G_n denote the empirical CDF of $\{X_i\}$ and $\{Y_i\}$. Then, the L_p norm between F_n and G_n satisfies:

1. if $\rho > 0$,
$$\mathbb{P}(\|F_n - G_n\|_p \geq \sqrt{\frac{\sigma}{n}}) \leq C_1 \exp\left(-\frac{1}{\sigma}\right); \tag{1}$$
2. if $\rho = 0$,
$$\mathbb{P}(\|F_n - G_n\|_p \leq \sqrt{\frac{\sigma}{n}}) \leq C_2 \exp\left(-\frac{1}{\sigma}\right), \tag{2}$$

for $p \in [1, \infty)$ with universal positive constants C_1 and C_2.

Proof Sketches: We prove the inequalities (1) and (2) separately.

For inequality (1), by Ding et al. (2021) (Need to figure out the specific derivation), we have

$$\|F_n - G_n\|_p \leq C\sqrt{\frac{\sigma}{n}},$$

for some positive constant C and general $p \geq 1$ when n is large enough. Therefore, for a fixed $q \geq 1$ and fixed $\lambda > 0$, we have

$$\begin{aligned}\mathbb{E}[\exp(\lambda\|F_n - G_n\|_q)] &= \mathbb{E}\left[1 + \sum_{k=1}^{\infty} \frac{\lambda^k \|F_n - G_n\|_q^k}{k!}\right] \\ &= 1 + \sum_{k=1}^{\infty} \frac{\lambda^k \mathbb{E}[\|F_n - G_n\|_q^k]}{k!} \\ &\leq 1 + \sum_{k=1}^{\infty} \frac{\lambda^k C^k (\sqrt{\sigma/n})^k}{k!} \\ &= \exp(\lambda C\sqrt{\sigma/n}),\end{aligned} \tag{3}$$

where the first and last equation follows by power series. The inequality (3) indicates that $\|F_n - G_n\|_q$ is a sub-Gaussian with sub-Gaussianity $\xi^2 \lesssim \sigma^2/n$; i.e.,

$$\mathbb{E}[\exp(\lambda\|F_n - G_n\|_q)] \lesssim \exp(\lambda C\sqrt{\sigma/n} + \lambda^2\xi^2/2).$$

Hence, by Chernoff bound, we have

$$\begin{aligned}\mathbb{P}(\|F_n - G_n\|_q \geq C\sqrt{\frac{\sigma}{n}} + t) &\leq \mathbb{P}(\|F_n - G_n\|_q \geq \mathbb{E}[\|F_n - G_n\|_q] + t) \\ &\leq \exp\left(-\frac{t^2}{2\xi^2}\right) \\ &\lesssim \exp\left(-\frac{nt^2}{\sigma^2}\right).\end{aligned}$$

Take $t \asymp \sqrt{\frac{\sigma}{n}}$. We have

$$\mathbb{P}(\|F_n - G_n\|_q \geq C'\sqrt{\frac{\sigma}{n}}) \leq C_1 \exp\left(-\frac{1}{\sigma}\right).$$

For inequality (2), we follow the inequality (14) in Ding et al. (2021) (Also need to figure out the specific derivation).

Therefore, when n is large enough, we have Conjecture 1.

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.