## Appendix

In this appendix, we discuss the convergence rate of $\sup_{f\in\mathcal{F}}\big|\mathrm{E}_N f-\mathrm{E}f\big|$ for an ID empirical process satisfying the conditions $(C1)-(C3)$.

Given a number $\widetilde{x}>1$, consider the following equation with respect to $\gamma>0$

$$\widetilde{x}-(\widetilde{x}+1)\ln(\widetilde{x}+1)=-\widetilde{x}^{\gamma}, \tag{44}$$

and denote its solution as

$$\gamma(\widetilde{x}):=\frac{\ln\big((\widetilde{x}+1)\ln(\widetilde{x}+1)-\widetilde{x}\big)}{\ln(\widetilde{x})}. \tag{45}$$

Then, we have for any $0<\widetilde{\gamma}\leq\gamma(\widetilde{x})$,

$$\widetilde{x}-(\widetilde{x}+1)\ln(\widetilde{x}+1)\leq-\widetilde{x}^{\widetilde{\gamma}}. \tag{46}$$

By using (46), we can represent the upper bound of $\sup_{f\in\mathcal{F}}\big|\mathrm{E}_N f-\mathrm{E}f\big|$ for the aforementioned ID empirical process as follows.

**Theorem 6.1** *Follow notations in Theorem 4.5 and assume that the conditions $(C1)$-$(C3)$ are all valid. Then, for any $\xi>0$ such that $N\xi^2\geq 32\max\{A^2,B^2\}$ and $\frac{\xi\beta_1 R}{4(\beta_1^2\beta_2K^2+V)}>1$, letting*

$$\begin{aligned}\epsilon=&2\exp\Big\{\Lambda_{\mathcal{F}}(2N)+\frac{N(\beta_1^2\beta_2K^2+V)}{\beta_1^2R^2}\\&\times\Gamma\left(\frac{\xi\beta_1R}{4(\beta_1^2\beta_2K^2+V)}\right)\Big\},\end{aligned} \tag{47}$$

*we have with probability at least $1-\epsilon$,*

$$\sup_{f\in\mathcal{F}}\big|\mathrm{E}_N f-\mathrm{E}f\big|\leq\left(\frac{4\beta_1R\big(\Lambda_{\mathcal{F}}(2N)-\ln(\epsilon/2)\big)}{N\left(\frac{\beta_1R}{4(\beta_1^2\beta_2K^2+V)}\right)^{\gamma-1}}\right)^{1/\gamma}, \tag{48}$$

*where $0<\gamma\leq\gamma\left(\frac{\xi\beta_1R}{4(\beta_1^2\beta_2K^2+V)}\right)$.*

**Proof.** It can be directly resulted from the combination of (19), (44), (45) and (46). ■

The above theorem shows that with probability at least $1-\epsilon$,

$$\sup_{f\in\mathcal{F}}\big|\mathrm{E}_N f-\mathrm{E}f\big|\leq O\left(\left(\frac{\Lambda_{\mathcal{F}}(2N)}{N}\right)^{\frac{1}{\gamma}}\right). \tag{49}$$

Furthermore, according to Lemma 4.2, if $VC(\mathcal{F})\leq D$, we have for any $N>D/2$,

$$\sup_{f\in\mathcal{F}}\big|\mathrm{E}_N f-\mathrm{E}f\big|\leq O\left(\left(\frac{\ln(2\mathrm{e}N/D)}{N/D}\right)^{\frac{1}{\gamma}}\right). \tag{50}$$

In order to find the convergence rate of $\sup_{f\in\mathcal{F}}\big|\mathrm{E}_N f-\mathrm{E}f\big|$, we have to study the upper bound of the function $\gamma(x)$ $(x>1)$. According to (45), for any $x>1$, we consider the derivative of $\gamma(x)$

$$\begin{aligned}\gamma'(x)=&\frac{\ln(x+1)}{\ln(x)\big((x+1)\ln(x+1)-x\big)}\\&-\frac{\ln\big((x+1)\ln(x+1)-x\big)}{x(\ln x)^2},\end{aligned} \tag{51}$$

and draw the function curve of $\gamma'(x)$ in Fig. 1.

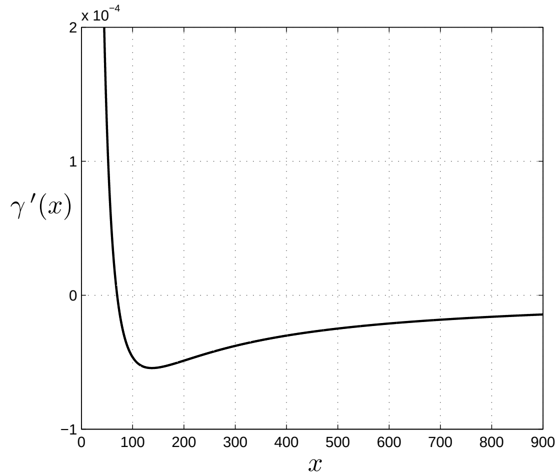


Figure 1: The Function Curve of $\gamma'(x)$

As shown in Fig. 1, there is only one solution to the equation $\gamma'(x)=0$ $(x>1)$. Letting the solution be $\widehat{x}$, we then have $\gamma'(x)>0$ $(1<x<\widehat{x})$ and $\gamma'(x)<0$ $(x>\widehat{x})$. Meanwhile, by (51), there holds that

$$\lim_{x\rightarrow+\infty}\gamma'(x)=0. \tag{52}$$

Therefore, we obtain that

$$\widehat{x}=\arg\max_{x>1}\gamma(x). \tag{53}$$

Our further numerical experiment shows that the value of $\widehat{x}$ approximately equals to 69.8517 and the maximum of $\gamma(x)$ $(x>1)$ is not larger than 1.3 (*cf.* Fig. 2). Thus, according to (49) and (50), we can obtain that with probability at least $1-\epsilon$,

$$\sup_{f\in\mathcal{F}}\big|\mathrm{E}_N f-\mathrm{E}f\big|\leq O\left(\left(\frac{\Lambda_{\mathcal{F}}(2N)}{N}\right)^{\frac{1}{1.3}}\right), \tag{54}$$

and

$$\sup_{f\in\mathcal{F}}\big|\mathrm{E}_N f-\mathrm{E}f\big|\leq O\left(\left(\frac{\ln(2\mathrm{e}N/D)}{N/D}\right)^{\frac{1}{1.3}}\right). \tag{55}$$

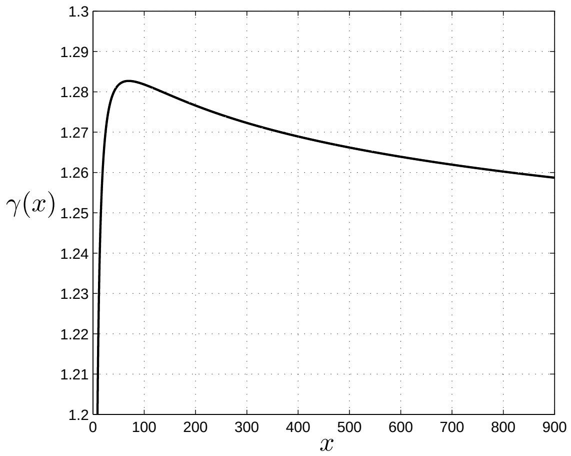


Figure 2: The Function Curve of $\gamma(x)$

Compared with the results of the generic i.i.d. empirical process given in (Vapnik, 1999)

$$\sup_{f\in\mathcal{F}}\big|\mathrm{E}_N f-\mathrm{E}f\big|\leq O\left(\left(\frac{\Lambda_{\mathcal{F}}(2N)}{N}\right)^{\frac{1}{2}}\right),$$

and

$$\sup_{f\in\mathcal{F}}\big|\mathrm{E}_N f-\mathrm{E}f\big|\leq O\left(\left(\frac{\ln(2\mathrm{e}N/D)}{N/D}\right)^{\frac{1}{2}}\right),$$

ID empirical process can provide a faster convergence rate of $\sup_{f\in\mathcal{F}}\big|\mathrm{E}_N f-\mathrm{E}f\big|$.