Detailed proof for the convergence

Loïc Fosse[1,2],
Philippe Formont[3,6,7,8],
Eric Aubinais[3,5],
Pablo Piantanida[3,4,6,7],
[1]Orange Research, Lannion, France
[2]CNRS, LIS, Aix Marseille Université, France
[3]CNRS, Université Paris-Saclay
[4]CNRS, CentraleSupélec
[5]Laboratoire de mathématiques d'Orsay (LMO), France
[6]International Laboratory on Learning Systems (ILLS), Canada
[7]Mila - Quebec AI Institute, Canada
[8]ÉTS Montreal, Canada

In this document, we provide a proof of the convergence of our deficiency estimator.

Theorem 1 (Convergence)**.** *If we suppose that for all $x \in \mathsf{X}$, $\mathbb{P}_{V|X=x}$ is absolutely continuous with respect to the Lebesgue measure, then we have,*

$$\hat{L}_n(f_m^*, K_m^*, x_m^*) \underset{n,m\to\infty}{\xrightarrow{\mathbb{P}}} L^*$$

Proof. First, we prove that $(f_m^*, K_m^*, x_m^*) \underset{m\to\infty}{\overset{w}{\rightsquigarrow}} (f^*, K^*, x^*)$ where (f^*, K^*, x^*) satisfies $L^* = L(f^*, K^*, x^*)$, and $\overset{w}{\rightsquigarrow}$ denotes the weak convergence. Let $\Lambda \subseteq \mathcal{M}(\mathsf{V} \mid \mathsf{U})$ be the set of kernels K such that $(f, x) \mapsto L(f, K, x)$ is continuous. Note that if K^* is such that $K^* \circ \mathbb{P}_{U|X}$ is not absolutely continuous with respect to the Lebesgue measure, then it is equivalent to $L^* = 1$. But there exist kernels K such that $K \circ \mathbb{P}_{U|X}$ is absolutely continuous with respect to the Lebesgue measure, implying that $K^* \in \Lambda$. For any $K \in \Lambda$, we let $(f_K^*, x_K^*) = \arg\max_{f,x} L(f, K, x)$, then $K \mapsto L(f_K^*, K, x_K^*)$ is continuous. Alternatively, for any $K \in \Lambda$, let $\mathbb{M}_m^K = \hat{L}_m(\cdot, K, \cdot)$ be an empirical process. Then we have $\mathbb{M}_m^K \underset{m\to\infty}{\overset{w}{\rightsquigarrow}} \mathbb{M}^K = L(\cdot, K, \cdot)$ in $\ell_c^\infty(\mathcal{F} \times \mathsf{X})$ as $\mathbb{M}^K$ is deterministic. By [1, Theorem 1], we have $(f_{m,K}^*, x_{m,K}^*) := \arg\max_{f,x} \mathbb{M}_m^K \overset{w}{\rightsquigarrow} (f_K^*, x_K^*)$. Now, it holds by the continuous mapping theorem [2] that

$$\left[\mathbb{G}_m : K \mapsto -\mathbb{M}_m^K(f_{m,K}^*, x_{m,K}^*)\right] \overset{w}{\rightsquigarrow} \left[\mathbb{G} : K \mapsto -\mathbb{M}^K(f_K^*, x_K^*)\right],$$

where $\mathbb{G}$ is deterministic. Now, as $\mathbb{G}$ is continuous on Λ, by [1, Theorem 1], it holds that $K_m^* = \arg\max_K \mathbb{G}_m(K) \underset{m\to\infty}{\overset{w}{\rightsquigarrow}} \arg\max_K \mathbb{G}(K) = K^*$. Therefore, given that (f^*, K^*, x^*) is deterministic, it holds that $(f_m^*, K_m^*, x_m^*) \underset{m\to\infty}{\overset{w}{\rightsquigarrow}} (f^*, K^*, x^*)$, which is equivalent to $L(f_m^*, K_m^*, x_m^*) \underset{m\to\infty}{\overset{\mathbb{P}}{\to}} L^*$. Letting $D(f, K\|w,t) = \mathbb{E}_{Z\sim K(\cdot|w,t)}[\mathbb{1}_{f(Z)=t}]$, the result follows with the following inequalities

$$\mathbb{P}\left(\left|L^* - \hat{L}_n(f_m^*, K_m^*, x_m^*)\right| \geqslant \varepsilon\right) \leqslant \mathbb{P}\left(\frac{1}{n}\sum_{(w,t)\sim\mathcal{T}_n^{x_m^*}} |L^* - D(f_m^*, K_m^*\|w,t)| \geqslant \varepsilon\right) \tag{1}$$

$$\leqslant \mathbb{E}_m\left[\mathbb{P}\left(\frac{1}{n}\sum_{(w,t)\sim\mathcal{T}_n^{x_m^*}} |L^* - \mathbb{E}_{w,t}\left[D(f_m^*, K_m^*\|w,t) \mid m\right]| \geqslant \varepsilon/2 \mid m\right)\right] \tag{2}$$

$$+ \mathbb{E}_m\left[\mathbb{P}\left(\frac{1}{n}\sum_{(w,t)\sim\mathcal{T}_n^{x_m^*}} |D(f_m^*, K_m^*\|w,t) - \mathbb{E}_{w,t}\left[D(f_m^*, K_m^*\|w,t) \mid m\right]| \geqslant \varepsilon/2 \mid m\right)\right] \tag{3}$$

$$\leqslant \mathbb{P}\left(|L^* - L(f_m^*, K_m^*, x_m^*)| \geqslant \varepsilon/2\right) + \mathbb{E}_m\left[\frac{Var(D\|m)}{n(\varepsilon/2)^2}\right] \tag{4}$$

$$\leqslant \mathbb{P}\left(|L^* - L(f_m^*, K_m^*, x_m^*)| \geqslant \varepsilon/2\right) + \frac{1}{n\varepsilon^2}, \tag{5}$$

where the third inequality is an application of Bienaymé-Tchebychev inequality, and the last inequality comes from the fact that $Var(D\|m) \leqslant 1/4$ a.s. because $D \in [0,1]$ a.s. $\square$

References

[1] Gregory Cox, "A generalized argmax theorem with applications," *arXiv preprint arXiv:2209.08793*, 2022.

[2] Henry B Mann and Abraham Wald, "On stochastic limit and order relationships," *The Annals of Mathematical Statistics*, 1943.