Conformal Risk Control for Ordinal Classification
(Supplementary Material)

Yunpeng Xu[1] **Wenge Guo**[*2] **Zhi Wei**[1]

[1]Department of Computer Science, New Jersey Institute of Technology, Newark, New Jersey, 07102, USA
[2]Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, New Jersey, 07102, USA

A A GENERALIZATION OF THEOREM 2 IN Angelopoulos et al. [2022]

In the Supplementary Material, we will present a generalization of Theorem 2 in Angelopoulos et al. [2022] on the lower bound of conformal risk control, which is needed in the proof of Theorem 1 in our paper. This result itself might be of independent interest to other applications. For convenience, we use the same notations as in Angelopoulos et al. [2022] in the following discussion.

Suppose that $\mathcal{C}_\lambda : \mathcal{X} \to 2^{\mathcal{Y}}$ is a given sequence of functions of an input $X \in \mathcal{X}$ that outputs a prediction set $\mathcal{C}(X) \subseteq \mathcal{Y}$, which is indexed by a threshold $\lambda \in \Lambda$, and $\mathcal{L}(Y, \mathcal{C}_\lambda(X)) \in (-\infty, B]$ be a given loss function of any observation (X, Y) and the corresponding prediction set $\mathcal{C}_\lambda(X)$. For the calibration observations $(X_i, Y_i)_{i=1}^n$ and the test observation (X_{n+1}, Y_{n+1}), let $L_i(\lambda) = L(Y_i, C_\lambda(X_i))$ for $i = 1, \ldots, n+1$ and $\hat{R}_n(\lambda) = (L_1(\lambda) + \ldots + L_n(\lambda))/n$. The value of λ is determined according to the following algorithm:

$$\hat{\lambda} = \inf\left\{\lambda : \frac{n}{n+1}\hat{R}_n(\lambda) + \frac{B}{n+1} \leq \alpha\right\},$$

where $\alpha \in (0, B)$ is the given desired risk level upper bound. Let $D = \{\lambda : J(\hat{R}_{n+1}, \lambda) > 0\}$ denote the set of discontinuities in $\hat{R}_{n+1}$, where $J(\mathcal{L}, \lambda)$ is the jump function defined below,

$$J(\mathcal{L}, \lambda) = \lim_{\epsilon \to 0+} \mathcal{L}(\lambda - \epsilon) - \mathcal{L}(\lambda),$$

which quantifies the size of the discontinuity in the loss function $\mathcal{L}$ at a point λ. For any $\lambda \in D$, define

$$s(\lambda) = |\{i : J(L_i, \lambda) > 0\}|,$$

the number of $L_i(\lambda)$ which are discontinuous at λ. Regarding $s(\lambda)$, we assume that

$$\sup_{\lambda \in \Lambda} s(\lambda) \leq M, \quad \text{almost surely},$$

where M is a non-negative integer. Specifically, if $M = 0$, this assumption implies that for any λ, $P(J(L_i, \lambda) > 0) = 0$ for $i = 1, \ldots, n+1$, which is the exactly original discontinuity assumption in Theorem 2 of Angelopoulos et al. [2022].

Under the above relaxed assumption, we generalize Theorem 2 in Angelopoulos et al. [2022] as follows. This result is applicable to the ordinal classification setting.

Theorem 4. *In the settings of Theorem 1 of Angelopoulos et al. [2022], further assume that L_i are i.i.d, $L_i > 0$ and*

$$\sup_{\lambda \in \Lambda} s(\lambda) \leq M, \quad \textit{almost surely},$$

[*]Corresponding author

Accepted for the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023).

where M is a non-negative integer. Then,

$$E[L_{n+1}(\hat{\lambda})] \geq \alpha - \frac{(M+2)B}{n+1}.$$

Specifically, if $M = 0$, the above result reduces to Theorem 2 in Angelopoulos et al. [2022]. To show Theorem 4, we need to generalize Lemma 1 of Angelopoulos et al. [2022] as follows and then use the similar arguments as in the proof of Theorem 2 therein along with this lemma.

Lemma 1. In the settings of Theorem 4, any jumps in the empirical risk are bounded, i.e.,

$$\sup_{\lambda \in \Lambda} J(\hat{R}_n, \lambda) \leq \frac{(M+1)B}{n}, \quad \text{almost surely.}$$

This lemma can be proved by using the similar arguments as in the proof of Lemma 1 of Angelopoulos et al. [2022].

References

Anastasios N. Angelopoulos, Stephen Bates, Adam Fisch, Lihua Lei, and Tal Schuster. Conformal risk control. *arXiv:2208.02814*, 2022.