1 The Effectiveness of Quantifier

Theorem 1. *The average miscoverage ratio of confidence intervals $\{\mathcal{C}_t\}_{t=1}^T$ produced by the Adaptive Conformal Uncertainty Quantifier will converge to α, i.e.*

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\prod_{j=1}^{h}(x_t^j \notin \mathcal{C}_t^j) \overset{a.s.}{=} \alpha.$$

Proof. We first demonstrate that for a fixed α, the coverage rate will be equal or less than α, and then further prove that for the adaptive version as using adaptive α_t instead of α, the actual miscoverage rate will gradually converge to α. The proof is inspired by that of conformal prediction in [Vovk et al.(2005), Sun and Yu(2024), Gibbs and Candes(2021)]. The following lemma shows the validity of the conformal predictive distributions, whose cumulative distribution function is constructed as Eqn. (1).

$$\hat{F}_j\left(s\right) := \frac{1}{|\mathcal{L}_{cal}|+1}\left(\tau + \sum_{\ell\in\mathcal{L}_{cal}}\mathbb{I}(l^j < s)\right) \tag{1}$$

Lemma 1 ([Vovk et al.(2017)])**.** *Given a loss function $\ell : \mathcal{Z}\to\mathbb{R}$ and a data sample $z\sim\mathcal{Z}$, calculate the loss as $l=\ell(z)$. Then, the cumulative distribution $\hat{F}_j(\cdot)$ constructed as Eqn. (1) is valid in the sense that $\mathbb{P}_{\mathcal{Z}}\left[\hat{F}_j\left(s_j\right)\le 1-\alpha\right] = 1-\alpha$, for any $0<\alpha<1$.*

Given a test data sample $z_t = (X_{t-L:t-1}, Y_{t:t+H-1}) \sim \mathcal{Z}$, we want to prove that the confidence intervals $C_t = \{c_t, \cdots, c_{t+H-1}\}$ output by ACCI satisfies:

$$\mathrm{P}\left[y_j \in C_j\right] \ge 1-\alpha, \quad \forall j\in\{t+1,\ldots,t+H\}$$

Define a partial order for k-dimensional vectors $\preceq$ as $\mathbf{u}\preceq\mathbf{v}$ i.f.f. $\forall j\in\{1,\ldots,h\}, \mathbf{u}_j\le\mathbf{v}_j$. For every data point in $\mathcal{L}_{cop}$, we evaluate the cumulative probability of the loss metric with the estimated conformal predictive distributions: $\mathcal{U}=\{\mathbf{u}^i\}_{i\in\mathcal{L}_{\mathrm{cop}}}$, $\mathbf{u}^i=(u_1^i,\ldots,u_H^i)=\left(\hat{F}_1\left(s_1^i\right),\ldots,\hat{F}_k\left(s_H^i\right)\right)$.

We define an empirical multivariate quantile function for $\mathcal{U}$, a set of k-dimensional vectors, based on the partial order:

$$Q(\mathcal{U};\alpha) \triangleq \inf_{\mathbf{u}^*}\left\{\mathbf{u}^* \;\middle|\; \left(\frac{1}{|\mathcal{U}|}\sum_{\mathbf{u}\in\mathcal{U}} sign(\mathbf{u}\preceq\mathbf{u}^*)\right)\ge 1-\alpha\right\}.$$

We first calculate $\mathbf{u}_j=\hat{F}_j\left(\ell\left(z_t\right)_j\right)$ for $j\in\{1,\ldots,h\}$. Let $\mathbf{u}^*=Q(\mathcal{U}\cup\{\infty\};\alpha), \mathbf{u}^*\in[0,1]^h$. An important observation for the conformal prediction proof is that if $\mathbf{u}^*\preceq\mathbf{u}_t$, then

$$Q(\mathcal{U}\cup\{\infty\};\alpha) = Q\left(\mathcal{U}\cup\{\mathbf{u}_t\};\alpha\right),$$

the quantile remains unchanged. This fact can be re-written as

$$\mathbf{u}_t\preceq Q(\mathcal{U}\cup\{\infty\};\alpha) \Longleftrightarrow \mathbf{u}_t\preceq Q\left(\mathcal{U}\cup\{\mathbf{u}_t\};\alpha\right)$$

The above describes the condition where $\mathbf{u}_t$ is among the $\lceil(1-\alpha)t\rceil$ smallest of $\mathcal{U}$. By exchangability, the probability of $\mathbf{u}_t$'s rank among $\mathcal{U}$ is uniform. Therefore,

$$\mathrm{P}\left[\mathbf{u}_t \preceq Q(\mathcal{U}\cup\{\infty\};\alpha)\right] = \frac{\lceil(1-\alpha)(|\mathcal{U}|+1)\rceil}{(|\mathcal{U}|+1)} \geq 1-\alpha \tag{2}$$

Note again that:

1. $\mathbf{u}^* = Q(\mathcal{U}\cup\{\infty\};\alpha) = \left(\hat{F}_1\left(s_1^*\right), \dots, \hat{F}_t\left(s_H^*\right)\right)$
2. $\mathbf{u}_t = \left(\hat{F}_1\left(s_t\right), \dots, \hat{F}_1\left(s_{t+H}\right)\right)$
3. The confidence intervals are constructed as:

$$C_j \leftarrow \left\{x : \|\mathbf{x}-\hat{\mathbf{x}}_j\| < s_j^*\right\} \tag{3}$$

By definition of $\preceq$, we have

$$\begin{aligned}\mathbf{u}_t \preceq \mathbf{u}^* &\Longleftrightarrow \forall j \in \{0,\dots,H-1\}, (\mathbf{u}_t)_j \leq \mathbf{u}_j^* \\ &\overset{\text{Lemma 1}}{\Longrightarrow} \forall j \in \{0,\dots,H-1\}, (s_t)_j \leq s_j^* \\ &\overset{\text{Eqn. (3)}}{\Longleftrightarrow} \forall j \in \{0,\dots,H-1\}, x_{t+j} \in C_j\end{aligned} \tag{4}$$

Combining Eqn. (2) and Eqn. (4), we have

$$\begin{aligned}\mathrm{P}\left[X_t \in C_t\right] &\geq \mathrm{P}\left[\mathbf{u}_t \preceq Q(\mathcal{U}\cup\{\infty;\alpha\})\right] \\ &\geq 1-\alpha\end{aligned} \tag{5}$$

It should be noted that the above proof is based on the assumption that the dataset is *exchangable*. Further, we discuss that replacing the fixed α with the adaptive miscoverage rate α_t, and updating it in an online manner will keep the validity of the conformal prediction without relying on this assumption.

Lemma 2. *With probability one we have that* $\forall t \in \mathbb{N}, \alpha_t \in [-\gamma, 1+\gamma]$.

Proof: Assume by contradiction that with positive probability $\{\alpha_t\}_{t\in\mathbb{N}}$ is such that $\inf_t \alpha_t < -\gamma$ (the case where $\sup_t \alpha_t > 1+\gamma$ is identical). Note that $\sup_t |\alpha_{t+1}-\alpha_t| = \sup_t \gamma\,|\alpha - \mathrm{err}_t| < \gamma$. Thus, with positive probability we may find $t \in \mathbb{N}$ such that $\alpha_t < 0$ and $\alpha_{t+1} < \alpha_t$. However,

$$\begin{aligned}\alpha_t < 0 &\Longrightarrow \hat{Q}_t\left(1-\alpha_t\right) = \infty \Longrightarrow \mathrm{err}_t = 0 \\ &\Longrightarrow \alpha_{t+1} = \alpha_t + \gamma\left(\alpha - \mathrm{err}_t\right) \geq \alpha_t\end{aligned}$$

and thus $\mathbb{P}\left(\exists t \text{ such that } \alpha_{t+1} < \alpha_t < 0\right) = 0$. We have reached a contradiction. With probability one we have that for all $T \in \mathbb{N}$,

$$\left|\frac{1}{T}\sum_{t=1}^{T} err_t - \alpha\right| \leq \frac{\max\left\{\alpha_1, 1-\alpha_1\right\}+\gamma}{T\gamma}.$$

In particular, $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathrm{err} \overset{\text{a.s.}}{=} \alpha$.

□

References

[Gibbs and Candes(2021)] Isaac Gibbs and Emmanuel Candes. 2021. Adaptive Conformal Inference Under Distribution Shift. *Advances in Neural Information Processing Systems (NeurIPS)* 34 (2021), 1660–1672.

[Sun and Yu(2024)] Sophia Sun and Rose Yu. 2024. Copula conformal prediction for multi-step time series forecasting. *International Conference on Learning Representations (ICLR)* (2024).

[Vovk et al.(2005)] Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. 2005. *Algorithmic learning in a random world*. Vol. 29. Springer.

[Vovk et al.(2017)] Vladimir Vovk, Jieli Shen, Valery Manokhin, and Min-ge Xie. 2017. Nonparametric predictive distributions based on conformal prediction. In *Conformal and probabilistic prediction and applications*. 82–102.