

Federated Conformal Prediction General

Min, Xia

May 5, 2024

1 Conformal Prediction General[1]

Definition 1.1 (Exchangeability). [3] For any r.v. x_1, \dots, x_k , we say they are exchangeable if for any permutation $\sigma : [k] \rightarrow [k]$ (bijection), $(x_1, \dots, x_k) \stackrel{d.}{=} (x_{\sigma(1)}, \dots, x_{\sigma(k)})$.

For conformal prediction two classes of targets are studied.

Definition 1.2 (Marginal Coverage). $(X, Y) \in \mathbb{R}^p \times \mathbb{R} \sim P_{XY}$ which is unknown. Given training set $Tr = \{(X_i, Y_i)\}_{i=1}^n$, and test on (X_{n+1}, Y_{n+1}) , both i.i.d.

C_α satisfies distribution-free marginal coverage at level $1 - \alpha$ if

$$P(Y_{n+1} \in C_\alpha(X_{n+1})) \geq 1 - \alpha, \quad \forall P_{XY}$$

The probability is with respect to $\{(X_i, Y_i)\}_{i=1}^{n+1}$.

Definition 1.3 (Conditional Coverage). $(X, Y) \in \mathbb{R}^p \times \mathbb{R} \sim P_{XY}$ which is unknown. Given training set $Tr = \{(X_i, Y_i)\}_{i=1}^n$, and test on (X_{n+1}, Y_{n+1}) , both i.i.d.

C_α satisfies distribution-free marginal coverage at level $1 - \alpha$ if

$$P\left(Y_{n+1} \in C_\alpha(X_{n+1}) \middle| X_{n+1} = x\right) \geq 1 - \alpha, \quad \forall P_{XY}$$

The probability is with respect to $\{(X_i, Y_i)\}_{i=1}^n$ and Y_{n+1} .

2 Standard Split Conformal Prediction

- First divide training set D into two sets: D_1 for proper training set and D_2 for calibration set. And let $n_i = |D_i|$, fit point predictor \hat{f}_1 on D_1 .
- Calculate residuals on D_2 : $R_i = |Y_i - \hat{f}_1(X_i)|$, $i \in D_2$.
- Find quantile on calibration residuals: $\hat{q}_2 = \lceil (1 - \alpha)(n_2 + 1) \rceil$ smallest of R_i , $i \in D_2$.
- Construct a conformal set: $C_\alpha(x) = [\hat{f}_1(x) - \hat{q}_2, \hat{f}_1(x) + \hat{q}_2]$.

Let $R_{n+1} = |Y_{n+1} - \hat{f}_1(X_{n+1})|$. Let rank statistic $R_{(j)}$ be the j -th smallest in R_i , $i \in D_2$, and $k_\alpha = \lceil (1 - \alpha)(n_2 + 1) \rceil$. As

$$\{Y_{n+1} \in C_\alpha(X_{n+1})\} = \{R_{n+1} \leq \hat{q}_2\} = \{R_{n+1} \leq R_{(k_\alpha)}\},$$

and R_i , $i \in D_2$, R_{n+1} are exchangeable, we have

$$\mathbb{P}\left(Y_{n+1} \in C_\alpha(X_{n+1}) \middle| D_1\right) \in \left[1 - \alpha, 1 - \alpha + \frac{1}{n_2 + 1}\right].$$

Assume a more general score function $V(x, y) = V((x, y); \hat{f}_1)$, define $R_i = V(X_i, Y_i)$ and change the conformal set to

$$C_\alpha(x) = \{y : V(x, y) \leq R_{(k_\alpha)}\}.$$

Remark 2.1. *Further condition on calibration set, which means conditioning on entire training set D and assume $R = V(x, y)$ has distribution F . As*

$$\{Y_{n+1} \in C_\alpha(X_{n+1})\} = \{R_{n+1} \leq R_{(k_\alpha)}\},$$

Assume the distribution function of $R_{(j)}$ is $F_{(j)}$, and we have

$$\begin{aligned} \mathbb{P}\left(\mathbb{P}\left(Y_{n+1} \in C_\alpha(X_{n+1}) \middle| D\right) \leq t\right) &= \mathbb{P}\left(\mathbb{P}\left(R_{n+1} \leq R_{(k_\alpha)} \middle| D\right) \leq t\right) \\ \text{condition on } D \text{ randomness comes from } R_{n+1}, &= \mathbb{P}\left(F(R_{(k_\alpha)}) \leq t\right) \\ &= \mathbb{P}\left(R_{(k_\alpha)} \leq F^{-1}(t)\right) \\ &= F_{(k_\alpha)}(F^{-1}(t)) \end{aligned} \tag{1}$$

rank statistic has density $F'_{(j)}(x) = jC_{n_2}^j x^{j-1}(1-x)^{n-j}f(x)$, thus take derivative on formula (1), and $\mathbb{P}\left(Y_{n+1} \in C_\alpha(X_{n+1}) \middle| D\right)$ has density

$$k_\alpha C_{n_2}^{k_\alpha} t^{k_\alpha-1} (1-t)^{n-k_\alpha}$$

3 Standard Full Conformal Prediction

Full CP has similar steps as split CP. It uses all data points for training.

- Fix any x and trial data y to construct training set $\{(X_1, Y_1), \dots, (X_n, Y_n), (x, y)\}$.
- Train point predictor \hat{f} on training set and define residuals $R_i = |Y_i - \hat{f}(X_i)|$, $i \in [n]$, $R_{n+1} = |y - \hat{f}(x)|$.
- Define j -th rank statistic of R_i , $i \in [n]$ as $R_{(j)}$, $k_\alpha = \lceil (1-\alpha)(n_2+1) \rceil$, and conformal set

$$C_\alpha(x) = \{y : R_{n+1} \leq R_{(k_\alpha)}\}$$

As $\{Y_{n+1} \in C_\alpha(X_{n+1})\} = \{R_{n+1} \leq R_{(k_\alpha)}\}$, and the exchangeability of data

$$\mathbb{P}(Y_{n+1} \in C_\alpha(X_{n+1})) \in \left[1 - \alpha, 1 - \alpha + \frac{1}{n+1}\right)$$

4 Federated Conformal Prediction Article1

Efficient Conformal Prediction under Data Heterogeneity[2]

Idea: The marginal coverage is measured over all training data and test points. However, if there is a high variability in the coverage probability as a function of the training data, the test coverage probability may be substantially below $1 - \alpha$ for a particular training set.

Definition 4.1 (empirical miscoverage rate). $\alpha(Tr) = P(Y_{n+1} \notin C_\alpha(X_{n+1}) \middle| Tr)$

References

- [1] Anastasios N Angelopoulos, Stephen Bates, et al. Conformal prediction: A gentle introduction. *Foundations and Trends® in Machine Learning*, 16(4):494–591, 2023.
- [2] Vincent Plassier, Nikita Kotelevskii, Aleksandr Rubashevskii, Fedor Noskov, Maksim Velikanov, Alexander Fishkov, Samuel Horvath, Martin Takac, Eric Moulines, and Maxim Panov. Efficient conformal prediction under data heterogeneity. In *International Conference on Artificial Intelligence and Statistics*, pages 4879–4887. PMLR, 2024.
- [3] Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(3), 2008.