Localized Conformal Prediction

Min

June 23, 2024

1 Conformal Prediction General[1]

Definition 1.1 (Exchangeability). *[7] For any r.v. $x_1, \cdots, x_k$, we say they are exchangeable if for any permutation $\sigma : [k] \to [k]$(bijection), $(x_1, \cdots, x_k) \overset{d.}{=} (x_{\sigma(1)}, \cdots, x_{\sigma(k)})$.*

Definition 1.2 (Weighted Exchangeability). *[8] For any r.v. $x_1, \cdots, x_k$, we say they are weighted exchangeable if their joint density canbe factorized as*

$$f(x_1, \cdots, x_k) = \prod_{i=1}^{k} w_i(x_i) \cdot g(x_1, \cdots, x_k),$$

where g is exchangeable, i.e., $g(x_1, \cdots, x_k) = g(x_{\sigma(1)}, \cdots, x_{\sigma(k)})$.

For conformal prediction two classes of targets are studied.

Definition 1.3 (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

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

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

Definition 1.4 (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

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

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

Definition 1.5 (Conformal Score Function)**.** *For data pair (X, Y) and point predictor and any loss function $V(\cdot,\cdot)$, call $R = S(X, Y) = V(Y, \hat{f}(X))$ be the conformal score(or residual).*

Definition 1.6 (Efficiency)**.** *X is some r.v. following the testing distribution and C_α is efficient if $\mathbb{E}\left[|C_\alpha(X)|\right]$ is small. Define Size$(C_\alpha) = \frac{1}{n}\sum_{i=1}^n |C_\alpha(X_i)|$.*

2 Localized CP Article1

Conformalized Quantile Regression [6]

The core idea is if conditional distribution function $F(y|X=x)$ is known and conditional quantile is $q_\alpha(x) = \inf\{y : F(y|X=x) \geq \alpha\}$, for $\alpha_1 = \alpha/2, \ \alpha_2 = 1-\alpha/2$ we can define conformal set to be $C_\alpha(x) = [q_{\alpha_1}(x), q_{\alpha_2}(x)]$. Next is to estimate quantiles from data.

Follow the split CP setting,

- 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 $\hat{q}_{\alpha_1}, \ \hat{q}_{\alpha_2}$ on D_1.
- Calculate cnformity scores on calibration set: $R_i = \max\{\hat{q}_{\alpha_1}(X_i) - Y_i, Y_i - \hat{q}_{\alpha_2}(X_i)\}$ for $i \in D_2$, and $R = \max\{\hat{q}_{\alpha_1}(X_i) - Y_i, Y_i - \hat{q}_{\alpha_2}(X_i)\}$
- Find the $\lceil(1-\alpha)(n_2+1)\rceil$-th empirical quantile of $R_i, \ i \in D_2$ as $\hat{q}$ and construct conformal set $C_\alpha(x) = [\hat{q}_{\alpha_1}(x) - \hat{q}, \hat{q}_{\alpha_2}(x) + \hat{q}]$

Note that $\{Y \in C_\alpha(X)\} = \{R \leq \hat{q}\}$. With exchangeability of $R_i, \ i \in D_2$ and R the coverage is assured.

Remark 2.1. *Can also define $R_1 = \hat{q}_{\alpha_1}(X_i) - Y_i, \ R_2 = Y_i - \hat{q}_{\alpha_2}(X_i)$ and their $\lceil(1-\alpha)(n_2+1)\rceil$-th empirical quantile $\hat{q}_1, \ \hat{q}_2$. Define conformal set $C_\alpha(X) = [\hat{q}_{\alpha_1}(X) - \hat{q}_1, \hat{q}_{\alpha_2}(X) + \hat{q}_2]$.*

3 Localized CP Article2

Distribution-Free Predictive Inference For Regression[5]

The setting is similar here. Consider the split setting, divide training set D into two sets: D_1 for proper training set and D_2 for calibration set.

- Train $\hat{f}(x)$ on D_1 as a point predictor and based on $(X_i, |Y_i - \hat{f}(X_i)|),\ i \in D_1$, train $\hat{\rho}$ as an estimator of conditional MAD $|Y - f(X)|\Big|X = x$. For a given test point X fix trial data y.
- Calculate scores on calibration set $R_i = \dfrac{|Y_i - \hat{f}(X_i)|}{\hat{\rho}(X_i)},\ i \in D_2,\ R = \dfrac{|y - \hat{f}(X)|}{\hat{\rho}(X)}$ and find the $\lceil(1-\alpha)(n_2+1)\rceil$-th empirical quantile $\hat{q}_\alpha$.
- Define conformal set $C_\alpha(X) = \{y : R \le \hat{q}_\alpha\}$.

Note that $\{Y \in C_\alpha\} = \{R \le \hat{q}_\alpha\}$ and $R,\ R_i,\ i \in D_2$ are exchangeable, it's easy to prove the coverage.

4 Localized CP Article3

Split Localized Conformal Prediction[3]

If score is defined by $R = |Y - \hat{f}(X)|$, the split CP follows setting $Y = \hat{f}(X) + \varepsilon$ where ε is independent of X. However this is not always true, we need to estimate the distribution of $R|X = x$.

Follow split CP setting, divide training set D into two sets: D_1 for proper training set and D_2 for calibration set.

We can estimate the distribution of $R|X = x$ with kernel smoothing. Assume distribution $F(R = r|X = x) = \mathbb{E}\mathbb{1}(R \le r|X = x)$, and NW estimator is

$$\hat{F}_h(R = r|X = x) = \sum_{i \in D_1} w_h(X_i|x)\mathbb{1}\{R_i \le r\},$$

where $w_h(X_i|x) = \dfrac{K(||g(X_i) - g(x)||/h)}{\sum\limits_{j \in D_1} K(||g(X_j) - g(x)||/h)}$ with some embedding function g. But directly find the α quantile of $\hat{F}_h(R = r|X = x)$ as $\hat{q}_\alpha$ and construct conformal set as

$\{y : R \leq \hat{q}_\alpha\}$ cannot guarantee coverage. Further, calculate a residual score on calibration set $R'_i = R_i - Q(\alpha, \hat{F}_h(R|X = X_i)),\ i \in D_2,\ R' = R - Q(\alpha, \hat{F}_h(R|X = X))$. The exchangeability still holds on R'_i. The conformal set is $C_\alpha = \{y : R \leq \hat{q}'_\alpha\}$, where $\hat{q}'_\alpha$ is $\lceil (1-\alpha)(n_2+1) \rceil$-th empirical quantile of $R'_i,\ i \in D_2$. Entire procedure is

- Train point predictor $\hat{f}$ on D_1 and calculate score $R_i = |Y_i - \hat{f}(X_i)|,\ i \in D_1$. Choose h to get NW estimator.
- Calculate on calibration set $R_i,\ Q(\alpha, \hat{F}_h(R|X = X_i)),\ i \in D_2$ and given X, for any trial data y, calculate R and $Q(\alpha, \hat{F}_h(R|X))$
- Calculate residual score on calibration set $R'_i = R_i - Q(\alpha, \hat{F}_h(R|X = X_i)),\ i \in D_2$ and $R' = R - Q(\alpha, \hat{F}_h(R|X))$. Find $\lceil (1-\alpha)(n_2+1) \rceil$-th empirical quantile of $R'_i,\ i \in D_2$
- Define conformal set $C_\alpha(X) = \{y : R' \leq \hat{q}'_\alpha\}$

The coverage guarantee comes from $\{Y \in C_\alpha(X)\} = \{R' \leq \hat{q}'_\alpha\}$ and the exchangeability within R' and $R'_i,\ i \in D_2$.

5 Localized CP Article4

Localized conformal prediction: a generalized inference framework for conformal prediction[2] Assume have $Z_1 = (X_1, Y_1), \cdots, Z_n = (X_n, Y_n)$ at hand and $Y|X$ varies across different X. A new observation X_{n+1} arrives and the conformal set of Y_{n+1} is required. The core idea is that for a given score function $S = V(X, Y)$ and require estimate a correct α quantile of $S|X$.

First a natural idea is construct $S|X$'s distribution. Assume any kernel function $K(x_1, x_2)$ and let $p_{i,j} = \dfrac{K(X_i, X_j)}{\sum_{l=1}^{n} K(X_i, X_l)}$. Conditional on X_i assign larger weight to S_j that has X_j near X_i. Let $F_i = \sum_{j=1}^{n} p_{i,j} \delta_{S_j}$ which is similar to the first idea of article3. Directly take the $1-\alpha$ quantile cannot guarantee coverage. Take $E_Z = \{\{Z_i\}_{i=1}^{n+1} = \{z_i\}_{i=1}^{n+1}\}$ and with all Z_i drawn i.i.d., $P(Z_{n+1} = z_i | E_Z) = 1/(n+1)$. Write s_i for each z_i. As E_Z means set equivalent, there exists some permutation $\sigma : [n+1] \to [n+1]$ s.t. $S_i = s_{\sigma(i)}$ condition on E_Z. To be specific, F_i is the distribution construct by $S_i,\ i = 1, \cdots, n+1$ and F_i' by $s_i,\ i = 1, \cdots, n+1$.

$$P(S_{n+1} \leq Q(\alpha'; F_{n+1})|E_Z) = \sum_{i=1}^{n+1} P(S_{n+1} = s_i|E_Z) \mathbb{1}\left\{S_{n+1} \leq Q(\alpha'; F_{n+1})|E_Z, S_{n+1} = s_i\right\}.$$

The most important thing here is $Q(\alpha'; F_{n+1})|(E_Z, S_{n+1} = s_i) = Q(\alpha'; F_i')|E_Z$. The order of $S_1, \cdots, S_n$ doesn't influence F_{n+1}. Thus,

$$P(S_{n+1} \leq Q(\alpha'; F_{n+1})|E_Z) = \sum_{i=1}^{n+1} P(S_{n+1} = s_i|E_Z) \mathbb{1}\left\{s_i \leq Q(\alpha'; F_i')|E_Z\right\}.$$

For any trial data $Y_{n+1} = y$, take s_i and F_i' be calculated based on sample data. Thus find α' that makes $\sum_{i=1}^{n+1} P(S_{n+1} = s_i|E_Z) \mathbb{1}\left\{s_i \leq Q(\alpha'; F_i')\right\} \geq 1 - \alpha$ then take expectation on E_Z and we have

$$P(S_{n+1} \leq Q(\alpha'; F_{n+1})) \geq 1 - \alpha.$$

Thus the entire process is as follows

- Fix trial data $Y_{n+1} = y$ and calculate score $s_1, \cdots, s_{n+1}$ and $F_1', \cdots, F_{n+1}'$ based on $Z_1, \cdots, Z_n, Z_{n+1}' = (X_{n+1}, y)$.
- Find α' that makes $\sum_{i=1}^{n+1} P(S_{n+1} = s_i|E_Z) \mathbb{1}\left\{s_i \leq Q(\alpha'; F_i')\right\} \geq 1 - \alpha$.
- Include y in conformal set if $s_{n+1} \leq Q(\alpha'; F_{n+1}')$. In conclusion $C_\alpha(X_{n+1}) = \{y : s_{n+1} \leq Q(\alpha'; F_{n+1}')\}$.

Remark 5.1. ******!!!!!*

For agent $1, \cdots, K$ each with distribution P^k we can try to find some kernel K to estimate the distance between agent $i,\ j$ and further construct the distribution or find α'.

6 Localized CP Article5

Conformal prediction with local weights: randomization enables robust guarantees[4]

Assume using training dataset $(X_1, Y_1), \cdots, (X_n, Y_n)$ and a test point (X_{n+1}, Y_{n+1}) all come from distribution $P = P_X \times P_{Y|X}$. Given a new X_{n+1}, sample $\tilde{X}$ based on X_{n+1} with density $H(X_{n+1}, \cdot)$ where H be some kernel function. Thus $X_{n+1}, \tilde{X}$ has joint density

$$P_X(X_{n+1})H(X_{n+1}, \tilde{X}).$$

Conditional on $\tilde{X}$(means be considered as a given constant) and density of X_{n+1} is propotional to $P_X(X_{n+1})H(X_{n+1}, \tilde{X})$ and finally the density ratio between X_{n+1} and $X_i, i \leq n$ is propotional to $H(X_{n+1}, \tilde{X})$.

This means conditional on $\tilde{X}$ and X_{n+1} has a covariate shift $H(X_{n+1}, \tilde{X})$ according to the training dataset. Construct empirical distribution

$$\tilde{F} = \sum_{i=1}^{n} \tilde{w}_i \delta_{S_i} + \tilde{w}_{n+1}\delta_\infty, \ \tilde{w}_i = \frac{H(X_i, \tilde{X})}{\sum_{j=1}^{n+1} H(X_j, \tilde{X})}.$$

The conformal set is $C_\alpha(X_{n+1}) = \{y : S_{n+1} \leq Q(1-\alpha, \tilde{F})\}$.

The coverage $\mathbb{P}(Y_{n+1} \in C_\alpha(X_{n+1})) \geq 1 - \alpha$ according to Tibshirani[8].

This method has certain local conditional coverage property. Start with a covariate shift setting.

Theorem 6.1. *Assume $X_1, \cdots, X_n \sim P_X, X_{n+1} \sim P_X \circ g, Y|X \sim P_{Y|X}$, where $P_X \circ g$ has density $\propto P_X(x)g(x)$. $g(x)$ is propotional to the density ratio. Assume $M = ||g||_\infty$ and $L_{g,2\epsilon,A} = \sup_{x,x' \in A, ||x-x'|| \leq 2\epsilon} |g(x) - g(x')|/2\epsilon$,*

$$\mathbb{P}\left(Y_{n+1} \in C_\alpha(X_{n+1})\right) \geq 1 - \alpha - \frac{\inf_{A \subset \mathcal{X}, \epsilon > 0} \left(\epsilon L_{g,2\epsilon,A} + M\left(\mathbb{P}\left(||X - \tilde{X}|| > \epsilon\right) + P_X(A^C)\right)\right)}{E_{P_X} g(X)}.$$

To see this, fixed $\tilde{X}$ which is generated with kernel $H(\cdot, X_{n+1})$ based on $X_{n+1} \sim P_X \circ g$. Assume $X'_{n+1}|\tilde{X} \sim P_X \circ H(\cdot, \tilde{X}), Y'_{n+1}|X'_{n+1} \sim P_{Y|X}$. Use the marginal coverage property

$$\mathbb{P}\left(Y'_{n+1} \in C_\alpha(X'_{n+1}) \middle| \tilde{X}\right) \geq 1 - \alpha. \tag{1}$$

A commonly used lemma is

Lemma 6.2. *Assume two random variables Z, Z', sharing common sample space $\mathcal{Z}$, have density $p(z)$ and $p'(z')$, the total variation is defined as $d_{TV}(Z, Z') = \int |p(z) - p'(z)|dz/2$. For any $A \subset \mathcal{Z}$,*

$$\Big|\mathbb{P}(Z \in A) - \mathbb{P}(Z' \in A)\Big| \leq d_{TV}(Z, Z').$$

Consider $\mathbb{P}\left(Y_{n+1} \in C_\alpha(X_{n+1})\right)$ according to formula 1,

$$\mathbb{P}\left(Y_{n+1} \in C_\alpha(X_{n+1}) \Big| \tilde{X}\right) \geq 1 - \alpha - d_{TV}(X_{n+1}, X'_{n+1}),$$

note $d_{TV}(X_{n+1}, X'_{n+1})$ is calculated conditional on $\tilde{X}$. Condition on $\tilde{X}$, X_{n+1} has density $P_X(x) \circ g(x) \circ H(x, \tilde{X})$, and X'_{n+1} has $P_X(x) \circ H(x, \tilde{X})$, to be specific

$$\frac{d(P_X \circ g \circ H(\cdot, \tilde{X}))(x)}{d(P_X \circ H(\cdot, \tilde{X}))(x)} = g(x) \frac{\int H(x', \tilde{X}) dP_X(x')}{\int g(x') H(x', \tilde{X}) dP_X(x')} = \frac{g(x)}{E_{X' \sim P_X \circ H(\cdot, \tilde{X})}\left[g(X')\right]},$$

thus as $d_{TV}(p, q) = \int |p(x) - q(x)|dx/2 = \int |p(x)/q(x) - 1|dQ(x)/2 = E_Q|p/q - 1|/2$, and the denominator of the last formula only depends on $\tilde{X}$ which is constant condition on $\tilde{X}$, we have

$$\begin{aligned}
d_{TV}(X_{n+1}, X'_{n+1}) &= \frac{1}{2} E_{X \sim P_X \circ H(\cdot, \tilde{X})} \left\{ \Big| \frac{g(X)}{E_{X' \sim P_X \circ H(\cdot, \tilde{X})}\left[g(X')\right]} - 1 \Big| \right\} \\
&= \frac{E_{X \sim P_X \circ H(\cdot, \tilde{X})} \left\{ g(X) - E_{X' \sim P_X \circ H(\cdot, \tilde{X})}\left[g(X')\right] \right\}}{2E_{X' \sim P_X \circ H(\cdot, \tilde{X})}\left[g(X')\right]} \\
\text{(Jensen's Inequality)} &\leq \frac{E_{X, X' \sim P_X \circ H(\cdot, \tilde{X})} \Big| g(X) - g(X') \Big|}{2E_{X' \sim P_X \circ H(\cdot, \tilde{X})}\left[g(X')\right]} \\
&= \frac{E_{P_X} H(X, \tilde{X})}{2E_{P_X} g(X) H(X, \tilde{X})} E_{X, X' \sim P_X \circ H(\cdot, \tilde{X})} \Big| g(X) - g(X') \Big|
\end{aligned}$$

Now take expectation on $\tilde{X}$ which has density $d\tilde{P}_{\tilde{X}}(x) = \dfrac{\int H(X, x) P_X(X) g(X) dX}{\int P_X(X) g(X) dX} dx$. Further assume $dP_{\tilde{X}}(x) = \int H(X, x) P_X(X) dX dx$, thus

$$\begin{aligned}
&E_{\tilde{X} \sim \tilde{P}_{\tilde{X}}} d_{TV}(X_{n+1}, X'_{n+1}) \\
\leq &\int \frac{E_{P_X} H(X, \tilde{X})}{2E_{P_X} g(X) H(X, \tilde{X})} \frac{\int H(X, \tilde{X}) P_X(X) g(X) dX}{\int P_X(X) g(X) dX} E_{X, X' \sim P_X \circ H(\cdot, \tilde{X})} \Big| g(X) - g(X') \Big| d\tilde{X} \\
= &\int \frac{E_{P_X} H(X, \tilde{X})}{2E_{P_X} g(X)} E_{X, X' \sim P_X \circ H(\cdot, \tilde{X})} \Big| g(X) - g(X') \Big| d\tilde{X} \\
= &\frac{E_{\tilde{X} \sim P_{\tilde{X}}} E_{X, X' \sim P_X \circ H(\cdot, \tilde{X})} \Big| g(X) - g(X') \Big|}{2E_{P_X} g(X)},
\end{aligned}$$

if $X, X' \in A, |X - X'| < 2\epsilon$, $|g(X) - g(X')| \leq 2\epsilon L_{g,2\epsilon,A}$, otherwise $|g(X) - g(X')| \leq 2M$. Write $E_A = \{X, X' \in A, |X - X'| < 2\epsilon\}$,

$$E_A^C \subset \left\{X \notin A \text{ or } X' \notin A \text{ or } |g(X) - g(\tilde{X})| > \epsilon \text{ or } |g(X') - g(\tilde{X})| > \epsilon\right\}.$$

Thus the probability over E_A^C is bounded, based on $X|\tilde{X} \sim P_{\tilde{X}}$ has distribution $P_X \circ H(\cdot, \tilde{X})$,

$$\begin{aligned}
E_{P_{\tilde{X}}}\mathbb{P}(E_A^C) &\leq 2E_{P_{\tilde{X}}}\left\{\mathbb{P}_{X\sim P_X\circ H(\cdot,\tilde{X})}\left[X \notin A \text{ or } |g(X) - g(\tilde{X})| > \epsilon\right]\right\} \\
&\leq 2E_{P_{\tilde{X}}}\left\{\mathbb{P}_{X\sim P_X\circ H(\cdot,\tilde{X})}[X \notin A] + \mathbb{P}_{X\sim P_X\circ H(\cdot,\tilde{X})}\left[|g(X) - g(\tilde{X})| > \epsilon\right]\right\} \\
&= 2\mathbb{P}_{X\sim P_X}(A^C) + 2\mathbb{P}_{X\sim P_X,\tilde{X}\sim P_{\tilde{X}}}\left(|g(X) - g(\tilde{X})| > \epsilon\right),
\end{aligned}$$

holds for any A, take inferior limit gives the result.

Next consider conditional coverage property.

Theorem 6.3. *Assume* $(X_1, Y_1), \cdots, (X_{n+1}, Y_{n+1}) \overset{i.i.d}{\sim} P$, *for any* $B \subset \mathcal{X}$, *it holds that*

$$\mathbb{P}\left(Y_{n+1} \in C_\alpha(X_{n+1})\middle|X_{n+1} \in B\right) \geq 1 - \alpha - \frac{\inf_{\epsilon>0}\left\{P_X(bd_{2\epsilon}(B)) + P_{X,\tilde{X}}\left(||X - \tilde{X}|| > \epsilon\right)\right\}}{P_X(B)},$$

where $bd_{2\epsilon}(B) = \left\{x \in B : \inf_{x' \in B^C} ||x - x'|| \leq 2\epsilon\right\}$.

This theorem is easy to prove using previous theorem. Take $g(x) = \mathbb{1}(x \in B)$ and $A = (bd_{2\epsilon}(B))^C$. For any $x, x' \in A$, either $x, x' \in B$ or $x, x' \in B^C$ which leads to $g(x) = g(x')$ and $L_{g,2\epsilon,A} = 0$. Also $E_{P_X}g(X) = P_X(B)$ and $M = ||g||_\infty = 1$.

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] Leying Guan. Localized conformal prediction: A generalized inference framework for conformal prediction. *Biometrika*, 110(1):33–50, 2023.

[3] Xing Han, Ziyang Tang, Joydeep Ghosh, and Qiang Liu. Split localized conformal prediction. *arXiv preprint arXiv:2206.13092*, 2022.

[4] Rohan Hore and Rina Foygel Barber. Conformal prediction with local weights: randomization enables local guarantees. *arXiv preprint arXiv:2310.07850*, 2023.

[5] Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018.

[6] Yaniv Romano, Evan Patterson, and Emmanuel Candes. Conformalized quantile regression. *Advances in neural information processing systems*, 32, 2019.

[7] Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(3), 2008.

[8] Ryan J Tibshirani, Rina Foygel Barber, Emmanuel Candes, and Aaditya Ramdas. Conformal prediction under covariate shift. *Advances in neural information processing systems*, 32, 2019.