A Few Observations on Sample-Conditional Coverage in Conformal Prediction

by Aryan Alavi

Based on the paper by John Duchi

February 2025

Topics

- Conformal Prediction
- X-conditioned Coverage
- Group and Weighted conditioned Coverage
- Sample Conditioned Coverage
- Bounding weighted-conditioned results

Conformal Prediction

- Predict Y for pairs coming from $(X, Y) \in \mathcal{X} \times \mathcal{Y}$

- We want to find confidence sets $C(x) \subset \mathcal{Y}$

Marginal Coverage Goal

- Given n samples $(X_i, Y_i)_{i=1}^n$

- Estimate confidence sets $\widehat{C}$ such that:

$$\mathbb{P}\left(Y_{n+1} \in \widehat{C}(X_{n+1})\right) \geq 1 - \alpha$$

Scoring Functions and Confidence Sets

- Assume scoring function $s : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$

- Define confidence sets $C_\tau(x) := \{y \mid s(x, y) \leq \tau\}$

Example: Regression

- $s(x, y) = |f(x) - y|$

-

$$C_\tau(x) = \{y \in \mathbb{R} \mid |y - f(x)| \leq \tau\} = [f(x) - \tau, f(x) +$$

Split Conformal Approach

- Assume similarity functions $S_i = s(X_i, Y_i)$
- Sort them by value $S_{(1)} \leq S_{(2)} \leq \cdots \leq S_{(n+1)}$

Split Conformal Approach - Continued

- We will have

$$\mathbb{P}\left(S_{n+1} > S_{(\lceil(1-\alpha)(n+1)\rceil)}\right) \leq \alpha$$

Split Conformal Approach - Solution

- Proven by Romano et al(2019), following threshold for similarity can be proposed $\widehat{\tau} := \mathsf{Quant}_{(1-\alpha)(1+1/n)}(S_1, \ldots, S_n)$
- $\mathbb{P}\left(S_{n+1} > \widehat{\tau}\right) \leq \alpha.$

Split Conformal Approach - Confidence Sets

- Our confidence sets will be

$$\widehat{C}(x) := \{y \in \mathcal{Y} \mid s(x, y) \leq \widehat{\tau}\}$$

- Satisfies

$$\mathbb{P}(Y_{n+1} \in \widehat{C}(X_{n+1})) = \mathbb{P}\left(s(X_{n+1}, Y_{n+1}) \leq \widehat{\tau}\right) = \mathbb{P}\left(S_{n+1} \leq \widehat{\tau}\right) \geq 1 - \alpha$$

X-conditional Coverage

- A bigger goal for confidence sets

$$\mathbb{P}(Y_{n+1} \in \widehat{C}_n(X_{n+1}) \mid X_{n+1} = x) \geq 1 - \alpha$$

X-conditional Coverage: Impossible

- Focusing on the case of $\mathcal{Y} = \mathbb{R}$, Vovk(2012)

- Lebesgue measure $\mathrm{Leb}(\widehat{C}(x))$ almost always infinite

Group Conditional Coverage

- For some groups $G \subset \mathcal{X}$, we want to have

$$\mathbb{P}(Y_{n+1} \in \widehat{C}(X_{n+1}) \mid X_{n+1} \in G) \geq 1 - \alpha.$$

- Barber et al. and Jung et al. have studied this problem

Generalized Notation

$$\mathbb{P}(Y \in C(x) \mid X = x) = 1 - \alpha$$

$$\mathbb{E}\left[w(X)\left(1\left\{Y \in \widehat{C}(X)\right\} - (1-\alpha)\right)\right] = 0$$

For all bounded w

Equal Constraints

$$\mathbb{E}\left[w(X)\left(\mathbb{P}(Y \in \widehat{C}(X) \mid X) - (1-\alpha)\right)\right] = 0$$

$$\mathbb{E}\left[\left|\mathbb{P}(Y \in \widehat{C}(X) \mid X) - (1-\alpha)\right|\right] = 0$$

Conditional Coverage - Interpretation

- The conditional coverage translates to:

$$\mathbb{E}\left[w(X)1\left\{Y \in \widehat{C}(X)\right\}\right] \geq (1-\alpha)\mathbb{E}[w(X)]$$

Group Coverage - Interpretation

- Should just be true for all sensitive groups weighting function

$$w(x) \;=\; 1\{x \in G\}$$

W-weighted Coverage

- Achieving $\mathcal{W}$-weighted $((1-\alpha), \epsilon)$ means

$$\left|\mathbb{E}\left[w(X)\left(1\{Y \in C(X)\} - (1-\alpha)\right)\right]\right| \leq \epsilon$$

For all $w \in \mathcal{W}$

W-weighted Coverage: Example

- For some vector v and feature mapping $\mathcal{X} \to \mathbb{R}^d$

$$\mathcal{W} = \{w \mid w(x) = \langle v, \phi(x) \rangle\}$$

Sample-Conditional Coverage

- Consider the samples $(X_i, Y_i)_{i=1}^n$

- Their empirical distribution P_n

Solve as before

$$\widehat{\tau}_n := \inf \{t \in \mathbb{R} \mid P_n(S \leq t) \geq 1 - \alpha\}$$

$$\Downarrow$$

$$\widehat{C}_n(x) := \{y \in \mathcal{Y} \mid s(x, y) \leq \widehat{\tau}_n\}$$

Sample-Conditional Coverage: Goal

- With some confidence over the choice of sample, give data about the original data distribution

Sample-Conditional Coverage - Result

- With probability of at least $1 - e^{-2n\gamma^2}$ over the choice of data, we will have

$$\mathbb{P}(Y_{n+1} \in \widehat{C}_n(X_{n+1}) \mid P_n) \geq 1 - \alpha - \gamma$$

Proof Insights

- Using simple Hoeffding bound

- Symmetrization, bounded difference, uniform convergence, ….

Corollary

- Assuming scores are distinct with probability 1, we will get that with probability at least $1 - 2e^{-2n\gamma^2}$ over samples,

$$1 - \alpha - \gamma \leq \mathbb{P}(Y_{n+1} \in \widehat{C}_n(X_{n+1}) \mid P_n) \leq 1 - \alpha + \frac{1}{n} + \gamma$$

Sample-Conditional Coverage: pre-defined confidence

- With probability at least $1-\delta$, we will have

$$1-\alpha-\gamma_n(\delta) \leq \mathbb{P}(Y_{n+1} \in \widehat{C}_n(X_{n+1}) \mid P_n)$$

Sample-Conditional Coverage: pre-defined confidence

- Where

$$\gamma_n(\delta) := \frac{4\log\frac{1}{\delta}}{3n} + \sqrt{\left(\frac{4}{3n}\log\frac{1}{\delta}\right)^2 + \frac{2\alpha(1-\alpha)}{n}\log\frac{1}{\delta}} \leq \frac{8\log\frac{1}{\delta}}{3n} + \sqrt{\frac{2\alpha(1-\alpha)}{n}\log\frac{1}{\delta}}$$

Corollary

- If scores have a density

$$1 - \alpha - \gamma_n(\delta) \leq \mathbb{P}(Y_{n+1} \in \widehat{C}_n(X_{n+1}) \mid P_n) \leq 1 - \alpha + \gamma_n(\delta)$$

Quantile Loss

- We define quantile loss as $\ell_\alpha(t) := \alpha\,[t]_+ + (1-\alpha)\,[-t]_+$

- The quantile $\mathsf{Quant}_{1-\alpha}(Y) := \inf\{t \mid \mathbb{P}(Y \le t) \ge 1-\alpha\}$ is the minimizer for the $L(t) := \mathbb{E}[\ell_\alpha(t-Y)]$.

- $\widehat{\theta}$ The empirical minimizer of the quantile loss

Theorem

- Assume the representation mode earlier, and n >= d
- Assume u to be any vector in the norm-2 unit ball with $\langle u, \phi(x) \rangle \geq 0$
- Let the maximum norm of norm of representation vectors
- $\widehat{h}(x) = \langle \widehat{\theta}, \phi(x) \rangle$

Theorem

- We define the confidence sets

$$\widehat{C}_n(x) := \left\{ y \in \mathcal{Y} \mid s(x,y) \leq \widehat{h}(x) \right\}$$

- The result states there exists a constant $c \leq 2 + \alpha/\sqrt{d}$

Theorem

- We will have with probability at least $1 - e^{-nt^2}$

$$\mathbb{E}\left[\langle u, \phi(X_{n+1})\rangle \left(1\left\{Y_{n+1} \in \widehat{C}_n(X_{n+1})\right\} - (1-\alpha)\right) \mid P_n\right] \geq -cb_\phi \left(\sqrt{\frac{d}{n}\log\frac{n}{d}} + t\right)$$

Theorem

- And if similarity values are distinct with probability 1, with same probability

$$\mathbb{E}\left[\langle u, \phi(X_{n+1})\rangle \left(1\left\{Y_{n+1} \in \widehat{C}_n(X_{n+1})\right\} - (1-\alpha)\right) \mid P_n\right] \leq 3b_\phi \left(\sqrt{\frac{d}{n}\log\frac{n}{d}} + t + \frac{d}{3n}\right)$$

Thanks for your attention!