Chap. 4: Conditional Coverage

Sunghee Park

Dept. of Statistics
Seoul National University

September 16, 2025

Table of contents

Contents

Conditional Coverage of the Conformal Prediction

- We have already guaranteed the marginal coverage of the conformal prediction.

$$1 - \alpha \leq \mathbb{P}(Y_{n+1} \in \mathcal{C}(X_{n+1}))$$

- This chapter focuses on the conditional coverage, which is $\mathbb{P}(Y_{n+1} \in \mathcal{C}(X_{n+1})|W)$ and satisfies the equation below:

$$\mathbb{E}[\mathbb{P}(Y_{n+1} \in \mathcal{C}(X_{n+1})|W)] = \mathbb{P}(Y_{n+1} \in \mathcal{C}(X_{n+1}))$$

- However, that relationship doesn't guarantee

$$1 - \alpha \leq \mathbb{P}(Y_{n+1} \in \mathcal{C}(X_{n+1})) \Rightarrow 1 - \alpha \leq \mathbb{P}(Y_{n+1} \in \mathcal{C}(X_{n+1})|W)$$

AT ALL.

What this chapter will cover

Recall that the conditional coverage is $\mathbb{P}(Y_{n+1} \in \mathcal{C}(X_{n+1})|W)$.
This chapter will cover

- Training-conditional coverage (W is related to $\mathcal{D}$)
- Test-conditional coverage (W is related to X_{n+1})
- and others...

Contents

Goals and Assumptions

The goal is to make $\mathbb{P}(Y_{n+1} \in \mathcal{C}(X_{n+1})|\mathcal{D}) > 1 - \alpha$ almost surely.

For some theoretical support of the conditional coverage, we need to make strong assumptions.

- We assume split conformal prediction. Which means that the score $s((x, y); \mathcal{D})$ and $\mathcal{D}$ are independent.
- The data points are i.i.d

If then, we can conclude Theorem 4.1 below.

Theorem 4.1. Distribution of the training conditional coverage

With previous assumptions,

$$\mathbb{P}\left(\mathbb{P}(Y_{n+1} \in \mathcal{C}(X_{n+1}) \mid \mathcal{D}_n\right) \leq 1 - \alpha - \Delta)$$
$$\leq F_{\text{Beta}((1-\alpha)(n+1), \alpha(n+1))}(1 - \alpha - \Delta) \leq e^{-2n\Delta^2},$$

Sketch of the proof of Theorem 4.1

- Let $s(X, Y) \sim F$ (CDF), $(X, Y) \sim P$, $S_i \sim F$ (i.i.d)
- By the algorihm of the CP, we can induce that

$$Y_{n+1} \in \mathcal{C}(X_{n+1}) \Leftrightarrow S_{n+1} \leq S_{(k)} \quad (k = \lceil (1-\alpha)(n+1) \rceil)$$

and

$$\mathbb{P}(Y_{n+1} \in \mathcal{C}(X_{n+1}) \mid \mathcal{D}_n) = \mathbb{P}(S_{n+1} \leq S_{(k)} \mid \mathcal{D}_n) = F(S_{(k)})$$

- Since $S_i \sim F$, by using the basic property of CDF (sim. as Probability integral transform), $F(S_i) \sim U_i$ and F preserves the order, (monotone increase) $F(S_{(i)}) \sim U_{(i)}$,
Let $U_i \sim U[0, 1]$ (i.i.d)

$$\begin{aligned} \mathbb{P}\big(F(S_{(k)}) \leq 1 - \alpha - \Delta\big) &= \mathbb{P}\big(U_{(k)} \leq 1 - \alpha - \Delta\big) \\ \leq \mathbb{P}\big(U^*_{(k)} \leq 1 - \alpha - \Delta\big) &= F_{\mathrm{Beta}((1-\alpha)(n+1),\, \alpha(n+1))}(1 - \alpha - \Delta) \end{aligned}$$

holds since the property of Beta distribution.

The meaning of Theorem 4.1

Theorem 4.1. Distribution of the training conditional coverage

With previous assumptions,

$$\mathbb{P}\left(\mathbb{P}(Y_{n+1} \in \mathcal{C}(X_{n+1}) \mid \mathcal{D}_n) \leq 1 - \alpha - \Delta\right)$$
$$\leq F_{\mathrm{Beta}((1-\alpha)(n+1),\, \alpha(n+1))}(1 - \alpha - \Delta) \leq e^{-2n\Delta^2},$$

- This theorem implies when we choose α' for CP, which is 'stricter' then α, we can make

$$\mathbb{P}\left(\mathbb{P}\left(Y_{n+1} \notin \mathcal{C}\left(X_{n+1}\right) \middle| \mathcal{D}_n\right) \leq \alpha\ \right) \geq 1 - \delta$$

- In detail, we can choose α' as:

$$F_{\mathrm{Beta}((1-\alpha')(n+1), \alpha'(n+1))}(1 - \alpha) = \delta$$

- It no longer guarantee in only the 'exchangeable' data (c.f. marginal case)

Hardness result for training-conditional coverage

Actually, it is impossible to fully guarantee training conditional coverage for full conformal prediction

Theorem 4.3. Hardness result for training-conditional coverage

Let P be any distribution on $\mathcal{X} \times \mathcal{Y}$ s.t. P_x is nonatomic*, there exists symmetric conformal score ftn s s.t. when running full conformal prediction with this choice of s,

$$\mathbb{P}\left(\mathbb{P}(Y_{n+1} \in \mathcal{C}(X_{n+1})|\mathcal{D}_n) = 0\right) \geq \alpha - O\left(\sqrt{\frac{\log n}{n}}\right)$$

where the probability is taken with respect to the training set $\mathcal{D}_n = ((X_1, Y_1), \ldots, (X_n, Y_n))$ drawn i.i.d. from P.

*nonatomic $\Leftrightarrow$ atom$(P) = \{z|\mathbb{P}_P(z) > 0\} = \emptyset$

Sketch of the proof of Theorem 4.3

- Constructive Proof
- Since P is nonatomic, $\exists a : \mathcal{X} \to \{0, 1, ..., n-1\}$ with $a(X)$ has a equal prob. in domain when $X \sim P$
- For dataset $\mathcal{D} = ((x_1, y_1), \ldots, (x_k, y_k))$ and an additional data point (x, y), Consider

$$s\big((x,y);\mathcal{D}\big) = \mathbb{1}\left\{\mathrm{mod}\left(-a(x) + \sum_{j=1}^{k} a(x_j), n\right) < N\right\}$$

- This score has quite useful properties in the proof:
  - Its domain is $\{0, 1\}$ - Simply think coverage fail only $S_{n+1} = 1 \wedge \hat{q} = 0 \Rightarrow \alpha_P(\mathcal{D}_n) = \mathbb{P}(S_{n+1} = 1, \hat{q} = 0 | \mathcal{D}_n)$
  - $S_{n+1} = s\big((X_{n+1}, Y_{n+1}); \mathcal{D}_{n+1}\big) = \mathbb{1}\left\{\mathrm{mod}\left(\sum_{i=1}^{n} a(X_i), n\right) < N\right\}$ (ftn unrelated to test point)
- Define $\mathcal{E}_{\mathrm{mod}}$ as the event that $\mathrm{mod}\left(\sum_{i=1}^{n} a(X_i), n\right) < N$.

Sketch of the proof of Theorem 4.3 (Cont')

- Let $\alpha_P(\mathcal{D}_n) := \mathbb{P}(Y_{n+1} \notin \mathcal{C}(X_{n+1}) \mid \mathcal{D}_n) = \mathbb{1}_{\mathcal{E}_{\text{mod}}} \cdot \mathbb{P}\left(\text{Quantile}\left(S_1, \ldots, S_{n+1}; 1-\alpha\right) = 0 \mid \mathcal{D}_n\right)$
- Using Sliding Window Method : Let $W_k = \left\{i \in \{0, \ldots, n-1\} : \text{mod}(-i+k-1, n) \geq N\right\}$, let $\mathcal{E}_{\text{unif}}$ be the event that

$$\sum_{i=1}^{n} \mathbb{1}\left\{a(X_i) \in W_k\right\} \geq (1-\alpha)(n+1) \text{ for all integers } k,$$

i.e., each window of indices W_k contains a sufficient fraction of the sample.

- $S_i = \mathbb{1}\left\{a(X_i) \notin W_{1+\sum_j a(X_j)}\right\}$ and by the property of $\mathcal{E}_{\text{unif}}$.

$$\mathbb{P}\left(\alpha_P(\mathcal{D}_n) = 1\right) \geq \mathbb{P}\left(\mathcal{E}_{\text{mod}} \cap \mathcal{E}_{\text{unif}}\right) \geq \mathbb{P}\left(\mathcal{E}_{\text{mod}}\right) - \mathbb{P}\left(\mathcal{E}_{\text{unif}}^c\right) = \frac{N}{n} - \mathbb{P}\left(\mathcal{E}_{\text{unif}}^c\right)$$

- Make upper bound of $\mathbb{P}(\mathcal{E}_{\text{mod}})$ using tail-prob. of Binomial dist. and set $N = \alpha n - O(n \log n)$

Contents

Goals

The goal is to make $\mathbb{P}(Y_{n+1} \in \mathcal{C}(X_{n+1})|X_{n+1}) > 1 - \alpha$ almost surely.

CP satisfying conditional coverage on discrete X

CP satisfying conditional coverage on discrete X

- Suppose that $\mathcal{X} = \{x_1, x_2, ..., x_{|\mathcal{X}|}\}$ then $\mathcal{C}(x_k) = \{y : s(x_k, y) \le \hat{q}_k\}$ where $\hat{q}_k = \mathrm{Quantile}\left((S_i)_{i\in[n], X_i = x_k}; (1-\alpha)(1+1/n_k)\right)$
- Similar to the naive CP, but now, we choose the quantile in a smaller group which has same X value.

- This will make the similar result to the naive CP when the dist. of $s(X, Y)|X = x_i$ are similar.
- The detailed proof will be discussed later.

CP satisfying conditional coverage on continuous X

- Actually, there is no "nice" prediction method which satisfying the conditional coverage on continuous X almost surely.

Theorem 4.4.

Suppose $\mathcal{C}$ is any procedure that satisfies distribution-free conditional coverage, i.e., for any distribution P on $\mathcal{X} \times \mathcal{Y}$, $\mathbb{P}\big(Y_{n+1} \in \mathcal{C}(X_{n+1}) \mid X_{n+1}\big) \geq 1 - \alpha$ holds almost surely, where the probability is taken with respect to $(X_1, Y_1), \cdots, (X_{n+1}, Y_{n+1}) \overset{\text{i.i.d.}}{\sim} P$.
Then, for any distribution P on $\mathcal{X} \times \mathcal{Y}$ for which the marginal P_X is nonatomic, $\mathbb{P}\big(y \in \mathcal{C}(x)\big) \geq 1 - \alpha$ for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$

Sketch of the proof of Theorem 4.4

- For all $(x, y) \in \mathcal{X} \times \mathcal{Y}$ and $\epsilon > 0$
- Using "Distridution-Free" feature of $\mathcal{C}$, Define a mixture dist. $P' = (1-\epsilon)P + \epsilon\delta_{(x,y)}$
- Since $\mathbb{P}_{P'}(X_{n+1} = x) > 0$, by assumption, $\mathbb{P}_{P'}\big(Y_{n+1} \in \mathcal{C}(X_{n+1}) \mid X_{n+1} = x\big) \geq 1 - \alpha$
- Moreover, P_X is nonatomic from the assumption, $\mathbb{P}_{P'}(Y_{n+1} = y | X_{n+1} = x) = 1$ and $\mathbb{P}_{P'}\big(y \in \mathcal{C}(X_{n+1}) \mid X_{n+1} = x\big) = \mathbb{P}_{P'}\big(y \in \mathcal{C}(x)\big) \geq 1 - \alpha$
- We can conclude $\mathbb{P}_{P}\big(y \in \mathcal{C}(x)\big) \geq \mathbb{P}_{P'}\big(y \in \mathcal{C}(x)\big) - d_{TV}(P^n, P'^n) \geq 1 - \alpha - n\epsilon$

CP satisfying conditional coverage on continuous X (Cont.)

Corollary 4.5. Lebesgue measure limitations of the continuous case.

Suppose $\mathcal{C}$ is any procedure that satisfies distribution-free conditional coverage, i.e., for any distribution P on $\mathcal{X} \times \mathcal{Y}$,

$$\mathbb{P}\big(\mathrm{Leb}(\mathcal{C}(x)) = \infty\big) \geq 1 - \alpha$$

where $\mathrm{Leb}(\,\cdot\,)$ denotes the Lebesgue measure.

Proof of Corollary 4.5.

- By definition of Lebesque measure,

$$\mathrm{Leb}(\mathcal{C}(x)) = \int_{\mathbb{R}} \mathbb{1}\{y \in \mathcal{C}(x)\}\, dy \le a \Longrightarrow \int_{y=0}^{a+b} \mathbb{1}\{y \in \mathcal{C}(x)\}\ \mathrm{d}y \le a$$
$$\Longleftrightarrow \int_{y=0}^{a+b} \mathbb{1}\{y \notin \mathcal{C}(x)\}\ \mathrm{d}y \ge b.$$

- Apply Markov's Inequality and Fubini's thm,

$$\mathbb{P}\left(\mathrm{Leb}(\mathcal{C}(x)) \le a\right) \le \mathbb{P}\left(\int_{y=0}^{a+b} \mathbb{1}\{y \notin \mathcal{C}(x)\}\ \mathrm{d}y \ge b\right)$$
$$\le \frac{\mathbb{E}\left[\int_{y=0}^{a+b} \mathbb{1}\{y \notin \mathcal{C}(x)\}\ \mathrm{d}y\right]}{b} = \frac{\int_{y=0}^{a+b} \mathbb{P}(y \notin \mathcal{C}(x))\, \mathrm{d}y}{b} \le \frac{(a+b)\alpha}{b}$$

- Take $b \to \infty$ and $a \to \infty$ in order

Contents

Relaxation Approach on test-conditional coverage

- So far, we concluded that it is impossible to achieve pointwise test-conditional coverage in continuous setting.
- One of the idea is relaxing the problem to discrete version - which is easily available.
- New goal: $1 - \alpha \leq \mathbb{P}(Y_{n+1} \in \mathcal{C}(X_{n+1}) | \mathcal{X}_k)$ for $k \in [K]]$ with $P(\mathcal{X}_k) > 0$

CP satisfying new goal (Binned Conditional Coverage)

- $\mathcal{C}(X_{n+1}) = \{y : s(X_{n+1}, y) \leq q_{k(\hat{x}_{n+1})}\}$ where $\hat{q}_k = \text{Quantile}\left((S_i)_{i \in [n], X_i \in \mathcal{X}_k}; (1-\alpha)(1+1/n_k)\right)$
- Similar to the naive CP, but now, we choose group-(score) quantile $\mathcal{X}_i$

Label-conditional coverage

- Goal : to make $\mathbb{P}(Y_{n+1} \in \mathcal{C}(X_{n+1}) | Y_{n+1}) > 1 - \alpha$ almost surely

CP satisfying the goal (Label-conditional coverage)

- $\mathcal{C}(X_{n+1}) = \{y : S_{n+1}^{y} \leq \hat{q}^{y}\}$ where $\hat{q}^{y} = \mathrm{Quantile}\left((S_i)_{i \in \mathcal{I}_y}; (1-\alpha)(1 + 1/|\mathcal{I}_y|)\right)$, $\mathcal{I}_y = \{i \in [n] : Y_i = y\}$
- Similar to the naive CP, but now, we group by y

Mondrian CP : Generalized CP

- Binned Conditional Coverage ($g(x, y) = k$ $(x \in \mathcal{X}_k)$) and Label-conditional coverage ($g(x, y) = y$) are the special cases.
- General Goals : to make $\mathbb{P}(Y_{n+1} \in \mathcal{C}(X_{n+1}) | g(X_{n+1}, Y_{n+1})) \geq 1 - \alpha$ w.p.1

Mondrian Conformal Prediction

- $\mathcal{C}(X_{n+1}) = \{y : S^y_{n+1} \leq \hat{q}^y\}$ where
  $\hat{q}^y = \mathrm{Quantile}\left((S_i)_{i \in \mathcal{I}_{g(X_{n+1}, y)}}; (1-\alpha)(1 + 1/|\mathcal{I}_{g(X_{n+1}, y)}|)\right)$,
  $\mathcal{I}_k = \{i \in [n] : g(X_i, Y_i) = k\}$
- CP with 'Grouping'

Proof of the validity of Mondrian CP

- First, we should show the statement below holds.

Lemma 4.7. Conditional exchangeability within a bin

Suppose $(X_1, Y_1), \ldots, (X_{n+1}, Y_{n+1})$ are exchangeable. Fix any subset $\mathcal{Z}_0 \subseteq \mathcal{X} \times \mathcal{Y}$, and for any fixed nonempty subset $I \subseteq [n+1]$, let $\mathcal{E}_I$ be the event that $\{i \in [n+1] : (X_i, Y_i) \in \mathcal{Z}_0\} = I$. If $\mathcal{E}_I$ has positive probability, then $((X_i, Y_i))_{i \in I}$ is exchangeable conditional on $\mathcal{E}_I$.

- The key of the proof is that for arbitrary $\sigma \in perm(I)$, think the extended permutation $\tilde{\sigma} \in perm([n+1])$ which satisfies the following equation and apply the exchangeablilty in $[n+1]$ wisely to show $\mathbb{P}((Z_i)_{i \in I} \in A,\ \mathcal{E}_I) = \mathbb{P}((Z_{\sigma(i)})_{i \in I} \in A,\ \mathcal{E}_I)$

$$\tilde{\sigma}(i) = \begin{cases} \sigma(i) & i \in I \\ i & i \notin I \end{cases}$$

Proof of the validity of Mondrian CP (Cont')

- Then, we prove the main theorem.
- First, by construction of the set $\mathcal{C}(X_{n+1})$, we can see that for any $k \in [K]$, on the event $g(X_{n+1}, Y_{n+1}) = k$,

$$Y_{n+1} \in \mathcal{C}(X_{n+1}) \iff \bar{p} := \frac{1 + \sum_{i \in [n], g(X_i, Y_i) = k} \mathbb{1}\{S_i \geq S_{n+1}\}}{1 + |\mathcal{I}_k|} > \alpha$$

- Next, fix any label $k \in [K]$ with $\mathbb{P}(g(X_{n+1}, Y_{n+1}) = k) > 0$, and let $\mathcal{Z}_0 = \{(x, y) \in \mathcal{X} \times \mathcal{Y} : g(x, y) = k\} \subseteq \mathcal{X} \times \mathcal{Y}$. By Lemma, the quantity

$$p = \frac{1 + \sum_{i \in [n]} \mathbb{1}\{(X_i, Y_i) \in \mathcal{Z}_0, S_i \geq S_{n+1}\}}{1 + \sum_{i \in [n]} \mathbb{1}\{(X_i, Y_i) \in \mathcal{Z}_0\}} = \bar{p}$$

satisfies $\mathbb{P}(p \leq \alpha \mid g(X_{n+1}, Y_{n+1}) = k) \leq \alpha$.
This completes the proof.

Another Relaxation on test-conditional coverage

- Another idea relaxing the problem is to weaken the condition of X.
- New goal: $\mathbb{P}\big(Y_{n+1} \in \mathcal{C}(X_{n+1}) \mid X_{n+1} \in \mathcal{X}_0\big) \geq 1 - \alpha$ for all P, and all $\mathcal{X}_0 \subseteq \mathcal{X}$ with $P_X(\mathcal{X}_0) \geq \delta$.
- Trivial Solution : much more strict target "Marginal coverage level" and use randomization

CP satisfying new goal

1. Construct $\mathcal{C}'(X_{n+1})$, using any method that guarantees marginal coverage at level $1 - c\alpha\delta$.
2. With probability $\frac{1-\alpha}{1-c\alpha}$, return $\mathcal{C}(X_{n+1}) = \mathcal{C}'(X_{n+1})$; otherwise, return $\mathcal{C}(X_{n+1}) = \varnothing$.

Prove of the previous method

- By construction of the method, we have

$$\mathbb{P}(Y_{n+1} \in \mathcal{C}(X_{n+1}) \mid X_{n+1} \in \mathcal{X}_0)$$
$$= \frac{1-\alpha}{1-c\alpha}\mathbb{P}(Y_{n+1} \in \mathcal{C}'(X_{n+1}) \mid X_{n+1} \in \mathcal{X}_0)$$

- Next

$$\mathbb{P}(Y_{n+1} \notin \mathcal{C}'(X_{n+1}) \mid X_{n+1} \in \mathcal{X}_0)$$
$$\leq \delta^{-1}\mathbb{P}(Y_{n+1} \notin \mathcal{C}'(X_{n+1}))\mathbb{1}_{X_{n+1}\in\mathcal{X}_0} \leq \delta^{-1} \cdot c\alpha\delta = c\alpha$$

Combining these two calculations proves the result.

Lower bound of the CI length in prev. method

Definition : L_P

let $L_P(1-t)$ be the minimum length of any *oracle* prediction interval $\mathcal{C}^P_{1-t}$ which is constructed given knowledge of the distribution P (i.e., not a distribution-free method), and has coverage level $1-t$:

$$L_P(1-t) = \inf\{\mathbb{E}_P[\mathrm{Leb}(\mathcal{C}^P_{1-t}(X))] \ : \\ \mathcal{C}^P_{1-t} \text{ satisfies } \mathbb{P}_P(Y \in \mathcal{C}^P_{1-t}(X)) \geq 1-t\}.$$

Theorem 4.13. Lower Bound of CI-length of the $\mathcal{C}$

Suppose $\mathcal{C}$ satisfies the distribution-free relaxed test-conditional coverage condition, and let $\mathcal{Y} = \mathbb{R}$. Then, for any distribution P on $\mathcal{X} \times \mathbb{R}$ for which the marginal P_X is nonatomic,

$$\mathbb{E}[\mathrm{Leb}(\mathcal{C}(X_{n+1}))] \geq \inf_{c\in[0,1]} \left\{ \frac{1-\alpha}{1-c\alpha} \cdot L_P(1-c\alpha\delta) \right\}.$$

Lower bound of the CI length in prev. method (Cont.)

- to prove the previous theorem, we need some lemmas.

Lemma 4.14

Suppose $\mathcal{C}$ satisfies the distribution-free relaxed test-conditional coverage condition. Let P be any distribution on $\mathcal{X} \times \mathcal{Y}$ s.t the marginal P_X is nonatomic, then
$\mathbb{P}_P\big(Y_{n+1} \in \mathcal{C}(X_{n+1}) \mid (X_{n+1}, Y_{n+1}) \in B\big) \geq 1 - \alpha$
for any $B \subseteq \mathcal{X} \times \mathcal{Y}$ with $P(B) \geq \delta$

Lower bound of the CI length in prev. method (Cont.)

- to prove the previous theorem, we need some lemmas.

Lemma 4.15. The sample–resample construction

Let P be a distribution on $\mathcal{Z}$, and let $m, M \geq 1$. Let P^m denote the corresponding product distribution on $\mathcal{Z}^m$—that is, the distribution of $(Z_1, \ldots, Z_m)$, where $Z_1, \ldots, Z_m \overset{\text{i.i.d.}}{\sim} P$. Moreover, let Q denote the distribution on $\mathcal{Z}^m$ obtained by the following process to generate $(Z_1, \ldots, Z_m)$:

1. Sample $Z^{(1)}, \ldots, Z^{(M)} \overset{\text{i.i.d.}}{\sim} P$, and define the empirical distribution $\widehat{P}_M = \frac{1}{M}\sum_{i=1}^{M} \delta_{Z^{(i)}}$;
2. Sample $Z_1, \ldots, Z_m \overset{\text{i.i.d.}}{\sim} \widehat{P}_M$.

Then

$$\mathrm{d}_{\mathrm{TV}}\left(P^m, Q\right) \leq \frac{m(m-1)}{2M},$$

where d_{TV} denotes the total variation distance between distributions.

Thank you!

References I