Extensions of Conformal Prediction

Heejoon Byun

Uncertainty Quantification Lab
Seoul National University

December 22, 2025

Table of contents

Contents

Beyond Miscoverage: Other Notions of Errors

Conformal prediction can be interpreted as controlling the expected miscoverage loss, $\mathbb{E}[\mathbb{1}\{Y_{n+1} \notin \mathcal{C}(X_{n+1})\}] = \mathbb{P}(Y_{n+1} \notin \mathcal{C}(X_{n+1}))$. However for some tasks, other notions of error can be more helpful:

1. **Coordinatewise coverage rate.**

$$\frac{1}{d}\sum_{j=1}^{d} \mathbb{1}\{(Y_{n+1})_j \notin \mathcal{C}(X_{n+1})_j\} \tag{1}$$

   where $Y_{n+1} \in \mathbb{R}^d$ and our prediction set is of the form $\mathcal{C}(X_{n+1}) = \mathcal{C}(X_{n+1})_1 \times \cdots \times \mathcal{C}(X_{n+1})_d$. (e.g. semantic segmentation)

2. **False negative rate.**

$$\frac{\sum_{j=1}^{d} \mathbb{1}\left\{(Y_{n+1})_j = 1, (\widehat{Y}_{n+1})_j = 0\right\}}{\sum_{j=1}^{d} \mathbb{1}\{(Y_{n+1})_j = 1\}} \tag{2}$$

   where $Y_{n+1} \in \{0,1\}^d$. (e.g. tumor segmentation)

Conformal Prediction (CP) vs. Conformal Risk Control (CRC)

	CP	CRC
Loss	$\mathbb{1}\left\{Y_{n+1} \notin \mathcal{C}_\lambda(X_{n+1})\right\} = \mathbb{1}\left\{s(X_{n+1}, Y_{n+1}) > \lambda\right\}$	$L(Y_{n+1}, \mathcal{C}_\lambda(X_{n+1}))$
Risk Estimate	$1 - \hat{F}(\lambda) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}\left\{s(X_i, Y_i) > \lambda\right\}$	$\hat{R}(\lambda) = \frac{1}{n}\sum_{i=1}^n L\left(Y_i, \mathcal{C}_\lambda(X_i)\right)$
Tuning	$\hat{\lambda} = \inf\left\{\lambda : 1 - \hat{F}(\lambda) \leq \alpha - (1-\alpha)/n\right\}$	$\hat{\lambda} = \inf\left\{\lambda : \hat{R}(\lambda) \leq \alpha - (1-\alpha)/n\right\}$
Guarantee	$\mathbb{P}(Y_{n+1} \notin \mathcal{C}_{\hat{\lambda}}(X_{n+1})) \leq \alpha$	$\mathbb{E}\left[L(Y_{n+1}, \mathcal{C}_{\hat{\lambda}}(X_{n+1}))\right] \leq \alpha$

Validity of Conformal Risk Control (CRC)

Theorem 1

Suppose that $(X_1, Y_1), \ldots, (X_{n+1}, Y_{n+1})$ are exchangeable, and that the loss L takes values in $[0, 1]$. Assume that the map $\lambda \mapsto L(y, \mathcal{C}_\lambda(x))$ is right-continuous and is monotone nonincreasing, for any (x, y). Then, with $\hat{\lambda}$ selected as in (5), we have

$$\mathbb{E}\left[L(Y_{n+1}, \mathcal{C}_{\hat{\lambda}}(X_{n+1}))\right] \leq \alpha. \tag{3}$$

Note that the monotonicity condition (i.e., requiring $\lambda \mapsto L(y, \mathcal{C}_\lambda(x))$ to be nonincreasing) is an immediate consequence of assuming nested sets $(\lambda_1 \leq \lambda_2 \implies \mathcal{C}_{\lambda_1}(x) \subseteq \mathcal{C}_{\lambda_2}(x))$, and monotonicity of the loss L $(\mathcal{C} \subseteq \mathcal{C}' \implies L(y, \mathcal{C}) \geq L(y, \mathcal{C}'))$.

We can also modify any general loss function $L(y, \mathcal{C}_\lambda(x))$ to $\widetilde{L}(y, \mathcal{C}_\lambda(x)) = \sup_{\lambda' \geq \lambda} L(y, \mathcal{C}_{\lambda'}(x))$, which is both nonincreasing and right continuous.

Proof Sketch

Proof

$$\lambda^* := \inf\left\{\lambda : \hat{R}(\lambda; \mathcal{D}_{n+1}) \leq \alpha\right\}. \quad (4)$$

$$\hat{\lambda} := \inf\left\{\lambda : \hat{R}(\lambda; \mathcal{D}_n) \leq \alpha - \frac{1-\alpha}{n}\right\}. \quad (5)$$

We can prove that

$$\mathbb{E}\left[L(Y_{n+1}, C_{\lambda^*}(X_{n+1}))\right] = \mathbb{E}\left[\hat{R}(\lambda^*; \mathcal{D}_{n+1})\right] \leq \alpha,$$

using exchangeability and right continuity, and that $\lambda^* \leq \hat{\lambda}$ using the fact that L takes values in $[0, 1]$. Therefore monotonicity of L implies that

$$\mathbb{E}\left[L(Y_{n+1}, C_{\hat{\lambda}}(X_{n+1}))\right] \leq \mathbb{E}\left[L(Y_{n+1}, C_{\lambda^*}(X_{n+1}))\right] \leq \alpha.$$

$\square$

Contents

Inference on Multiple Test Points

Goal. Use conformal prediction on multiple test points so that joint validity is guaranteed.

Family-Wise Error Rate (FWER) Control. Suppose we are given a set of test points $X'_1, \ldots, X'_m$.

- **FWER**: probability of *at least one* miscoverage among our m prediction sets.
- **FWER control**:

$$\mathbb{P}\big(Y'_1 \in \mathcal{C}(X'_1) \text{ and } \ldots \text{ and } Y'_m \in \mathcal{C}(X'_m)\big) \geq 1 - \alpha_{\text{FWER}}, \tag{6}$$

Asymptotic FWER

Proposition 2

Define the split conformal prediction set as

$$\mathcal{C}_{\hat{q}_n}(x) = \{y : s(x, y) \leq \hat{q}_n\}$$

$$\hat{q}_n = \mathrm{Quantile}\big(S_1, \dots, S_n; (1-\alpha)(1+1/n)\big)$$

Define $q^* = \mathrm{Quantile}(F, 1-\alpha)$, *where* F *is the CDF of* $s(X, Y)$ *when* $(X, Y) \sim P$. *Assume that* $F(q)$ *is continuous at* $q = q^*$, *and that* q^* *is the unique* $(1-\alpha)$*-quantile. Then*

$$\lim_{n \to \infty} \mathbb{P}\big(Y'_1 \in \mathcal{C}_{\hat{q}_n}(X'_1) \text{ and } \dots \text{ and } Y'_m \in \mathcal{C}_{\hat{q}_n}(X'_m)\big) = (1-\alpha)^m. \quad (7)$$

Note that this proposition asserts that the aim of FWER control will only hold (asymptotically) if we choose

$$1 - \alpha_{\mathsf{FWER}} = (1-\alpha)^m \iff \alpha = 1 - (1 - \alpha_{\mathsf{FWER}})^{1/m} \approx \alpha_{\mathsf{FWER}}/m,$$

i.e., this is nearly the Bonferroni correction.

Proof Sketch

Proof

We have

$$\mathbb{P}\left(S_1' \leq \hat{q}_n \text{ and } \ldots \text{ and } S_m' \leq \hat{q}_n\right) = \mathbb{E}\left[F(\hat{q}_n)^m\right] \to F(q^*)^m = (1-\alpha)^m.$$

since by Theorem 5.6, it holds almost surely that $\hat{q}_n \to q^*$. The rest are just applications of the continuous mapping theorem and dominated convergence theorem. □

Non-Asymptotic FWER

Proposition 3

In the setting of Proposition 2, for any fixed $n \geq 1$,

$$\mathbb{P}\left(Y_1' \in \mathcal{C}_{\hat{q}_n}(X_1') \text{ and } \ldots \text{ and } Y_m' \in \mathcal{C}_{\hat{q}_n}(X_m')\right) \geq (1-\alpha)^m. \quad (8)$$

Proof

From the proof of Proposition 2 and Jensen's inequality, we have that $\mathbb{P}\left(S_1' \leq \hat{q}_n \text{ and } \ldots \text{ and } S_m' \leq \hat{q}_n\right) = \mathbb{E}[F(\hat{q}_n)^m] \geq \mathbb{E}[F(\hat{q}_n)]^m$. Also, by the tower law,

$$\mathbb{E}[F(\hat{q}_n)] = \mathbb{E}[\mathbb{P}(Y_1' \in \mathcal{C}_{\hat{q}_n}(X_1') \mid \hat{q}_n)] = \mathbb{P}(Y_1' \in \mathcal{C}_{\hat{q}_n}(X_1')) \geq 1-\alpha,$$

where the last step holds by the distribution-free marginal coverage guarantee for the split conformal prediction method. □

These results show that a Bonferroni-type correction is unavoidable if we wish to obtain FWER control over a collection of test points.

Q1

- **Q.** 10.2.1 절에서 테스트 데이터가 여러개면 Bonferroni correction 처럼 전체 데이터에 대한 커버리지가 $1-\alpha$ 이상이 되게 하려면 임계값 q를 계산할때 α/m로 사용해야한다라고 이해했는데, 실제로 이렇게 많이 하나요?
- **A.** 잘 모르겠습니다. 일반적인 다중예측 상황에서 Bonferroni-type correction들은 대체로 매우 보수적이기에, 한번에 만들어야 하는 예측구간의 수가 많은 유전학이나 뇌과학, 신약 후보 발견 같은 분야에서는 FWER control 대신 FCR (false coverage rate) control로 목적을 바꾸는 식으로 이 문제를 회피하는 반면, 실수 하나의 대가가 매우 큰 임상실험 결과 예측 같은 분야나 재현성이 중요한 사회과학이나 입자물리 같은 분야는 FWER control을 사용하며, conformal prediction 기반 알고리즘 역시 대체로 이런 기조를 따를 것으로 생각됩니다. 찾아보니 자율주행에서의 경로예측과 같이 실수 하나하나의 대가가 매우 큰 분야에서는 보수적이더라도 FWER control이 되는 conformal prediction 기반 알고리즘들에 관심을 가지는 것 같습니다. (Lindemann et al., Safe Planning in Dynamic Environments Using Conformal Prediction, 2023 참고)

Contents

Outlier Detection

Goal. Use conformal prediction to detect outliers that are not drawn from P, the distribution of the calibration points. (i.e., simultaneously test for $H_{0,i} : (X'_i, Y'_i) \sim P$. for each test point $i \in [m]$)

Fals Discovery Rate (FDR) Control. Suppose we are given a set of test points $(X'_1, Y'_1), \ldots, (X'_m, Y'_m)$ (note that this time the responses are also observed).

- **FDR**: the expected proportion of false rejections;

$$\text{FDR} = \mathbb{E}\left[\frac{|\{i \in \mathcal{R} : H_{0,i} \text{ is true}\}|}{\max\{1, |\mathcal{R}|\}}\right], \tag{9}$$

  where $\mathcal{R} \subseteq \{1, \ldots, m\}$ is the set of *rejections* among the null hypotheses $H_{0,1}, \ldots, H_{0,m}$ being tested.

- **FDR control**:

$$\mathbb{E}\left[\frac{|\{i \in \mathcal{R} : H_{0,i} \text{ is true}\}|}{\max\{1, |\mathcal{R}|\}}\right] \leq \alpha_{\text{FDR}}. \tag{10}$$

Benjamini-Hochberg Algorithm & The PRDS Condition

The Benjamini–Hochberg algorithm is defined as

$$\mathcal{R} = \left\{ i : p_i \leq \frac{\alpha_{\mathrm{FDR}} \hat{k}}{m} \right\} \text{ where } \hat{k} = \max \left\{ k : \sum_{i=1}^{m} \mathbb{1} \left\{ p_i \leq \frac{\alpha_{\mathrm{FDR}} k}{n} \right\} \geq k \right\}, \tag{11}$$

where $\alpha_{\mathrm{FDR}} \in [0, 1]$ is the target FDR level. When applied to independent p-values, this algorithm has the property that FDR $\leq \alpha_{\mathrm{FDR}}$.

Even though the conformal p-values p_i defined above are not independent, they have a particular form of dependence that validates the Benjamini–Hochberg algorithm:

Definition 4 (Positive regression dependence on a subset (PRDS))

A random vector $W \in \mathbb{R}^m$ is PRDS on a set $I_0 \subseteq \{1, \ldots, m\}$ if for any $i \in I_0$ and any nondecreasing set A, [$(a_1, \ldots, a_m) \in A$ implies that any $(b_1, \ldots, b_m)$, with $b_i \geq a_i$ for all $i \in [m]$, must also lie in A.] $\mathbb{P}(W \in A \mid W_i = w_i)$ is nondecreasing in w_i.

Q2

- **Q.** target FDR level을 달성하기 위해 Benjamini-Hochberg algorithm을 도입하였는데, 이 알고리즘의 직관이 궁금합니다. 또한, 독립과 PRDS의 차이를 직관적으로 알고 싶습니다.
- **A.**

$$\mathcal{R} = \{i : p_i \leq \hat{t}\} \text{ where } \hat{t} = \max\left\{t : \frac{mt}{1 \vee \sum_{i=1}^{m} \mathbb{1}\{p_i \leq t\}} \leq \alpha_{\text{FDR}}\right\}, \tag{12}$$

The PRDS condition is equivalent to the following: For any $i \in I_0$ and any nondecreasing function f, $\mathbb{E}[f(W) \mid W_i = w_i]$ is nondecreasing in w_i (the equivalence follows from approximations with nondecreasing simple functions). This alternative statement implies that the W_i's are nonnegatively correlated with each other.

Conformal p-values are PRDS

Theorem 5 (Conformal p-values are PRDS)

Suppose $(X_1, Y_1), \ldots, (X_n, Y_n), (X'_1, Y'_1), \ldots, (X'_m, Y'_m)$ are independent. Assume also that $(X_i, Y_i) \sim P$ for all $i \in [n]$, and that $(X'_i, Y'_i) \sim P$ for all $i \in I_0$. Let s be a pretrained score function, and assume that all calibration and test scores are distinct almost surely. Then, the conformal p-values are PRDS on the subset I_0.

Corollary 6 (FDR control with conformal p-values)

In the setting above, let $\mathcal{R}$ be the set of rejections of the Benjamini–Hochberg algorithm, when run at target FDR level α_{FDR} based on the p-values $p_1, \ldots, p_m$. Then, the FDR is controlled:

$$\mathbb{E}\left[\frac{|\{i \in \mathcal{R} : H_{0,i} \text{ is true}\}|}{\max\{1, |\mathcal{R}|\}}\right] \leq \alpha_{\text{FDR}}. \tag{13}$$

Proof Sketch

Proof

Fix the first test point in I_0. First, for each $i \in \{2, \dots, m\}$, define a count $C_i = \sum_{j=1}^{n} \mathbb{1}\{s(X'_i, Y'_i) \leq s(X_j, Y_j)\} + \mathbb{1}\{s(X'_i, Y'_i) \leq s(X'_1, Y'_1)\}$.
Then, we can prove that $p_1 \perp\!\!\!\perp (C_2, \dots, C_m)$. Also for each $i \in \{2, \dots, m\}$, $s(X'_i, Y'_i) \leq s(X'_1, Y'_1) \iff C_i \geq (n+1)p_1$.
Now let $A \subseteq \mathbb{R}^m$ be an nondecreasing set. We need to verify that

$$t \mapsto \mathbb{P}((p_1, \dots, p_m) \in A \mid p_1 = t) = \mathbb{P}(V_t \in A \mid p_1 = t)$$

is nondecreasing in t where we define the random vector V_t as

$$V_t = \left(t, \frac{1 + C_2 - \mathbb{1}\{C_2 \geq (n+1)t\}}{n+1}, \dots, \frac{1 + C_m - \mathbb{1}\{C_m \geq (n+1)t\}}{n+1}\right).$$

Since we have proved that $p_1 \perp\!\!\!\perp (C_2, \dots, C_m)$ while V_t is a function of $(C_2, \dots, C_m)$, we have $\mathbb{P}(V_t \in A \mid p_1 = t) = \mathbb{P}(V_t \in A)$, which is a nondecreasing function, as desired. $\square$

Contents

Settings

- **Goal.** Guarantee coverage conditional on the outcome of a (data dependent) selection rule
- A selection rule $\mathcal{I} : (\mathcal{X} \times \mathcal{Y})^{n+1} \to 2^{[n+1]}$ is *symmetric* if the ordering of the data points does not affect which data points are selected. i.e.,

$$\sigma(i) \in \mathcal{I}(\mathcal{D}) \iff i \in \mathcal{I}(\mathcal{D}_\sigma), \tag{14}$$

  for any dataset $\mathcal{D}$, any permutation σ, and any index $i \in [n+1]$.
- The selected data are exchangeable conditional on the output of a symmetric selection rule:

Lemma 7 (Conditional exchangeability after selection)

Suppose $(X_1, Y_1), ..., (X_{n+1}, Y_{n+1})$ are exchangeable, and let $\mathcal{I}$ be a symmetric selection rule as in (14). Let $\mathcal{E}_I$ be the event that $\mathcal{I}(\mathcal{D}_{n+1}) = I$, for some fixed nonempty subset $I \subseteq [n+1]$. Assume $\mathcal{E}_I$ has positive probability. Then $((X_i, Y_i))_{i \in I}$ is exchangeable conditional on $\mathcal{E}_I$.

Selective Conformal Prediction

Theorem 8 (Coverage guarantee of selective conformal prediction)

Suppose $(X_1, Y_1), \ldots, (X_{n+1}, Y_{n+1})$ are exchangeable. Let s be a symmetric score function, and let $\mathcal{I}$ be a symmetric selection rule as in (14). Assume the event $n+1 \in \mathcal{I}(\mathcal{D}_{n+1})$ (i.e., the event that the test point is selected) has positive probability. Then the prediction set $\mathcal{C}(X_{n+1})$ defined as

$$\mathcal{C}(X_{n+1}) = \{y : S^y_{n+1} \le \hat{q}^y, n+1 \in \mathcal{I}(\mathcal{D}^y_{n+1})\}, \tag{15}$$

where

$$\hat{q}^y = \mathrm{Quantile}\left((S^y_i)_{i \in \mathcal{I}^y}; (1-\alpha)(1+1/|\mathcal{I}^y|)\right) \tag{16}$$

and $\mathcal{I}^y = \{i \in [n] : i \in \mathcal{I}(\mathcal{D}^y_{n+1})\}$ is the subset of training points that are selected when the selection rule is run with test point (X_{n+1}, y), satisfies

$$\mathbb{P}\left(Y_{n+1} \in \mathcal{C}(X_{n+1}) \mid n+1 \in \mathcal{I}(\mathcal{D}_{n+1})\right) \ge 1-\alpha. \tag{17}$$

Conformal p-Values & Selection

Corollary 9

In the setting of Lemma 7, define

$$p = \frac{1 + \sum_{i \in [n]} \mathbb{1}\left\{i \in \mathcal{I}(\mathcal{D}_{n+1}), S_i \geq S_{n+1}\right\}}{1 + \sum_{i \in [n]} \mathbb{1}\left\{i \in \mathcal{I}(\mathcal{D}_{n+1})\right\}}$$

where $S_i = s((X_i, Y_i); \mathcal{D}_{n+1})$ *for some symmetric score function* s*. If* $\mathbb{P}(n+1 \in \mathcal{I}(\mathcal{D}_{n+1})) > 0$*, then*

$$\mathbb{P}(p \leq \alpha \mid n+1 \in \mathcal{I}(\mathcal{D}_{n+1})) \leq \alpha$$

for any $\alpha \in [0, 1]$*.*

Q3

- **Q.** Selective coverage에서 $\mathcal{I}$가 symmetric하다는 가정이 들어가있는데, 만약 없다면 어떻게 될까요? 좀더 완화한 가정으로 selective CP를 할 수 있을까요?
- **A.** Selective CP의 selection-conditional coverage를 보장하기 위해서는 Lemma 7의 selection-conditional exchangeability가 필요하고, 이것이 성립하기 위해서는 selection rule이 반드시 symmetric해야 합니다. 따라서 이 가정이 없다면 더 이상 selective cp의 selection-conditional coverage는 보장할 수 없습니다. 따라서 이를 해결하기 위해서는 다른 알고리즘을 사용해야 하는데, 제 생각에는 selection 자체를 distribution shift로 해석하면 6단원의 weighted cp 알고리즘을 적용할 수 있을 것 같습니다.

Q4

- **Q.** mondrian CP의 확장 버전이 selective CP라고 나와있는데, 그러면 selective CP가 해결하는 mondrian CP의 단점들이 있나요? 있다면 무엇인가요?
- **A.** Mondrian CP는

$$\mathbb{P}(Y_{n+1} \in \mathcal{C}(X_{n+1}) \mid g(X_{n+1}, Y_{n+1}) = k) \geq 1 - \alpha \tag{18}$$

와 같은 coverage guarantee를 가지고, 따라서 Mondrian CP 만으로도 하나의 data point에만 의존하는 selection rule (e.g. $\hat{f}(X_i) \geq c$)에 대해서는 selection-conditional coverage를 보장할 수 있습니다. 이 결과를 calibration set에 속한 다른 여러 data point 에도 의존하는 selection rule (e.g. 위의 예에서 $\hat{f}$의 적합에 calibration data도 사용된 경우, top-k selection)로 확장시킨 것이 selective CP라고 이해하시면 되겠습니다.

Contents

Ensemble for Prediction Sets?

- **Q.** Can we combine multiple prediction sets so that the coverage guarantee is (almost) preserved?
- **A.** Yes. Aggregation via majority vote will preserve the coverage guarantee up to a factor of two:

Theorem 10 (Coverage guarantee for majority vote aggregation)

Suppose $\mathbb{P}(Y_{n+1} \in \mathcal{C}_k(X_{n+1})) \geq 1 - \alpha$ *for each* $k = 1, \ldots, K$. *Then the majority vote set*

$$\mathcal{C}^{\mathsf{mv}}(X_{n+1}) = \left\{ y : \frac{1}{K} \sum_{k=1}^{K} \mathbb{1}\left\{ y \in \mathcal{C}_k(X_{n+1}) \right\} > 1/2 \right\}. \tag{19}$$

satisties

$$\mathbb{P}\Big(Y_{n+1} \in \mathcal{C}^{\mathsf{mv}}(X_{n+1}) \Big) \geq 1 - 2\alpha. \tag{20}$$

This result follows almost immediately from Markov's inequality.

Post-Aggregation Calibration

Theorem 11 (Coverage guarantee for post-aggregation calibration)

Suppose $(X_1, Y_1), \ldots, (X_{n+1}, Y_{n+1})$ are exchangeable, and are independent of the pretrained prediction sets $\mathcal{C}_1, \ldots, \mathcal{C}_K$. Define

$$\mathcal{C}'_k(x;\beta) = \left\{y : \text{ for all } \epsilon > 0,\ y \in \mathcal{C}_k(x;\beta') \text{ for some } \beta' \le \beta + \epsilon\right\}, \tag{21}$$

$$\mathcal{C}^{\mathsf{mv}}(x;\lambda) = \left\{y : \frac{1}{K}\sum_{k=1}^{K} \mathbb{1}\left\{y \in \mathcal{C}'_k(x_{n+1};\lambda)\right\} > 1/2\right\}, \tag{22}$$

$$\hat{\lambda} = \inf\left\{\lambda \in [0,1] : \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{Y_i \notin \mathcal{C}^{\mathsf{mv}}(X_i;\lambda)\right\} \le \alpha - (1-\alpha)/n\right\}. \tag{23}$$

Then under the notation and definitions above,

$$\mathbb{P}\Big(Y_{n+1} \in \mathcal{C}^{\mathsf{mv}}(X_{n+1};\hat{\lambda})\Big) \ge 1 - \alpha. \tag{24}$$

Proof Sketch

Proof

Define a score function

$$s(x, y) = \inf \{\lambda : y \in \mathcal{C}^{\mathsf{mv}}(x; \lambda)\}.$$

Since $\lambda \mapsto \mathbb{1}\{y \in \mathcal{C}'_k(x; \lambda)\}$ is monotone nondecreasing and is right-continuous, we have that $\mathcal{C}^{\mathsf{mv}}(x; \lambda) = \{y : s(x, y) \leq \lambda\}$. By definition of $\hat{\lambda}$, we then have

$$\begin{aligned}\hat{\lambda} &= \inf \left\{ \lambda \in [0, 1] : \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{s(X_i, Y_i) \leq \lambda\} \geq (1 - \alpha)(1 + 1/n) \right\} \\ &= \mathrm{Quantile}\big(s(X_1, Y_1), \ldots, s(X_n, Y_n); (1 - \alpha)(1 + 1/n)\big).\end{aligned}$$

We can then see that $\mathcal{C}^{\mathsf{mv}}(X_{n+1}; \hat{\lambda})$ is exactly equal to the split conformal prediction set, $\{y : s(X_{n+1}, y) \leq \hat{\lambda}\}$, which then guarantees coverage. $\square$

Q5

- **Q.** Aggregating Conformal Set에서 해당하는 내용은 Conformal Prediction의 성질을 거의 사용하지 않는 것 같은데, 그러면 다른 모델에 대한 CI를 갖다 써도 문제가 없지 않을까요?
- **A.** 네, 맞습니다. 첫 번째 결과는 marginal coverage를 만족하는 prediction set이면 모두 성립하고, 두 번째 결과는 심지어 marginal coverage를 만족하지 않는 임의의 prediction set에 대해 성립합니다. 그러나 실제로 알고리즘을 적용할 때는 prediction set $\mathcal{C}_k$들의 monotonic right-continuous version인 $\mathcal{C}'_k$이 경우에 따라 계산이 어렵거나 불가능할 수 있으니 현실적으로는 처음부터 이러한 좋은 성질들을 가진 prediction set을 사용해야 할 것 같습니다.

Thank you!