Biometrika (2022), **xx**, x, *p.* 1
Printed in Great Britain

Supplements to *Localized Conformal Prediction: A Generalized Inference Framework for Conformal Prediction*

BY LEYING GUAN
Dept. of Biostatistics, Yale University,
New Haven, CT, USA.
leying.guan@yale.edu

In this Supplement, we describe a few supplemental Lemmas used in our proofs to results in the main paper in Appendix A. We then give proofs to results in the main paper in Appendix B and proofs to the supplemental Lemmas in Appendix C. We provide details of the construction of the dissimilarity measure $d(.,.)$ used in this paper and the automatic choice of h in Appendix D.

Without loss of generality, we always assume that $V_1 \leq V_2 \leq \ldots \leq V_n$ in this Supplement. We use $\mathbb{E}[.]$ to represent the expectation of a given variable and we refer to conformal prediction as CP and localized conformal prediction as LCP for convenience.

A. A COLLECTION OF SUPPLEMENTAL LEMMAS

Lemma A.1 describes the elementary relationship used in the proof from previous work on weighted conformal prediction (Barber et al., 2019), and we state it here for the reader's convenience. Lemma A.2 states the monotone dependence of $Q(\tilde{\alpha}; \hat{\mathcal{F}}_i(v))$ on $\tilde{\alpha}$ or v. Lemma A.3 is a core Lemma on the marginal coverage guarantee for LCP with strategically chosen $\tilde{\alpha}$. Lemma A.4 collects basic bounds used in the proofs of Theorem 3.

Lemma A.1. For any α and sequence $\{V_1, \ldots, V_{n+1}\}$, we have

$$V_{n+1} \leq Q(\alpha; \sum_{i=1}^{n} w_i \delta_{V_i} + w_{n+1}\delta_{V_{n+1}}) \Leftrightarrow V_{n+1} \leq Q(\alpha; \sum_{i=1}^{n} w_i \delta_{V_i} + w_{n+1}\delta_{\infty}),$$

where $\sum_{i=1}^{n} w_i \delta_{V_i} + w_{n+1}\delta_{V_{n+1}}$ and $\sum_{i=1}^{n} w_i \delta_{V_i} + w_{n+1}\delta_{\infty}$ are some weighted empirical distributions with weights $w_i \geq 0$ and $\sum_{i=1}^{n+1} w_i = 1$.

Lemma A.2. Suppose $\{V_i, i = 1, \ldots, n\}$, the target level α, and empirical weights p^H_{ij} are given. Then,

(i) Given V_{n+1}, $Q\left(\tilde{\alpha}; \hat{\mathcal{F}}_i(V_{n+1})\right)$ for $i = 1, \ldots, n+1$ and $Q\left(\tilde{\alpha}; \hat{\mathcal{F}}\right)$ are non-decreasing, right-continuous and piece-wise constant on $\tilde{\alpha}$, and with value changing only at the cumulative probabilities at different V_i.

(ii) Given $\tilde{\alpha}$, $Q\left(\tilde{\alpha}; \hat{\mathcal{F}}_i(v)\right)$ is non-decreasing on v for $i = 1, \ldots, n+1$.

(iii) If $V_{n+1} = v$ is accepted in the $C_V(X_{n+1})$ in Lemma 1, then v' is accepted for any $v' \leq v$.

Lemma A.3. Let $V_i = V(Z_i; \mathcal{Z})$ be the score for sample i, and Z_i is i.i.d generated for $i = 1, \ldots, n+1$. For any event

$$\mathcal{T} := \{\{Z_i, i = 1, \ldots, n+1\} = \{z_i := (x_i, y_i), i = 1, \ldots, n+1\}\},$$

we have

$$\mathbb{P}\{V_{n+1} \leq Q(\tilde{\alpha}; \sum_{i=1}^{n+1} p^H_{n+1,i}\delta_{V_i})|\mathcal{T}\} = \mathbb{E}\left[\frac{1}{n+1}\sum_{i=1}^{n+1} \mathbb{1}_{v_i \leq v^*_i}|\mathcal{T}\right],$$

© 2020 Biometrika Trust

where $v_i = V(z_i; (z_1, \dots, z_n, z_{n+1}))$, $v_i^* = Q(\tilde{\alpha}; \sum_{j=1}^{n+1} p_{i,j}^H \delta_{v_j})$ for $i = 1, 2, \dots, n+1$, and $\tilde{\alpha}$ can be random but is independent of the data conditional on $\mathcal{T}$. The expectation on the right side is taken over the randomness of $\tilde{\alpha}$ conditional on $\mathcal{T}$.

Lemma A.4. Suppose that Assumption 1 holds and $V(.)$ is a fixed function. For any x_0, define $B(x_0) = \sum_{j=1}^{n+1} H(x_0, X_i)$, $\Delta(x_0, X) = H(x_0, X) \max_v |P_{V|X}(v) - P_{V|x_0}(v)|$ and $\Delta(x_0) = \sum_{i=1}^{n} \Delta(x_0, X_i)$. Then,

(i) There exists a constant $C > 0$ such that, for all $x_0 \in [0,1]^p$, we have

$$\mathbb{P}\left\{B(x_0) \le \frac{nh_n^\beta}{2eL}\right\} \le \exp(-\frac{nh_n^\beta}{8L}), \qquad \frac{\Delta(x_0)}{B(x_0) \vee (nh_n^\beta)} \le Ch_n \ln(h_n^{-1}).$$

(ii) Set $B_i = B(X_i)$, $R_i = \frac{\sum_{j \ne i}\left(\mathbb{1}_{V_j < V_i} - P_{V|X_j}(V_i)\right)}{B_i}$. Then, for all V_i and $i = 1, \dots, n+1$, we have

$$\mathbb{P}\left\{|R_i| \ge \sqrt{\frac{\ln n}{B_i}} | \mathcal{X}, V_i\right\} \le \frac{2}{n^2}.$$

B. Proofs Propositions, Lemmas and Theorems

In this section, we provide proofs omitted from the main paper. We first give arguments to Proposition 1 and Proposition 2 for the counterexamples. We then present proofs to Theorem 1, Theorem 2, Lemma 1, Lemma 2 that characterize the marginal behavior of LCP and our implementation. After that, we prove Theorem 3 - 4 on the asymptotic and local behaviors of LCP-type procedures.

Proofs of the counter examples

B.1. *Proof of Proposition 1*

Proof. When $\sum_{i=1}^{n+1} H(X_{n+1}, X_i) < \frac{1}{1-\alpha}$, by definition, we have

$$\sum_{i=1}^{n} p_{n+1,i}^H = \frac{\sum_{i=1}^{n} H_{n+1,i}}{\sum_{i=1}^{n+1} H_{n+1,i}} < \frac{\frac{1}{1-\alpha} - 1}{\frac{1}{1-\alpha}} = \alpha.$$

We thus have $Q(\alpha; \hat{\mathcal{F}}) = \infty$, and consequently,

$$\mathbb{P}(Q(\alpha; \hat{\mathcal{F}}) = \infty) = \mathbb{P}(\sum_{i=1}^{n} p_{n+1,i}^H < \alpha) = \mathbb{P}(\sum_{i=1}^{n+1} H(X_{n+1}, X_i) < \frac{1}{1-\alpha}) \ge \varepsilon.$$
$$\mathbb{P}(Y_{n+1} \in C(X_{n+1})) \ge \mathbb{P}(Q(\alpha; \hat{\mathcal{F}}) = \infty) \ge \varepsilon.$$

B.2. *Proof of Proposition 2*

Proof. For $X_{n+1} \in \{\pm e_j, j = 1, \dots, p\}$, let n_0 is the number of samples with $X_i = 0$ and n_1 is the number of samples with $X_i = X_{n+1}$. The achieved conditional coverage at $\tilde{\alpha} = \alpha$ given $\mathcal{X} = X_{1:(n+1)}$ can be upper bounded as below:

$$\mathbb{P}\left\{V_{n+1} \le Q(\alpha; \hat{\mathcal{F}}) | \mathcal{X}\right\} = \mathbb{P}\left\{V_{n+1} \le Q(\alpha; \frac{1}{n_1 + n_0 + 1} \sum_{i: X_i \in \{0, X_{n+1}\}} \delta_{V_i} + \frac{1}{n_1 + n_0 + 1} \delta_\infty) \Big| \mathcal{X}\right\}$$
$$\overset{(a)}{\le} \frac{1}{n_1 + 1} + \frac{n_0 + n_1 + 1}{n_1 + 1}[\alpha - \frac{n_0}{n_0 + n_1 + 1}]_+, \tag{B.1}$$

Step (a) holds because:

– When $\alpha \le \frac{n_0}{n_1 + n_0 + 1}$, the α quantile of the weighted empirical distribution is 0, and we will have 0 coverage for $X_{n+1} \ne 0$ and (B.1) is true.

– When $\alpha > \frac{n_0}{n_1+n_0+1}$, the α quantile of the weighted empirical distribution in (B.1) is the $\lceil (n_1+n_0+1)\alpha \rceil - n_0$ largest value in $\{V_i : X_i = X_{n+1}\} \cup V_\infty$, which is the $\frac{\lceil (n_1+n_0+1)\alpha \rceil - n_0}{n_1+1}$ quantile of the unweighted empirical distribution formed by $\{V_i : X_i = X_{n+1}\} \cup V_\infty$. By Lemma A.1, $\{V_{n+1} \le Q(t; \{V_i : X_i = X_{n+1}\} \cup V_\infty)\} \Leftrightarrow \{V_{n+1} \le Q(t; \{V_i : X_i = X_{n+1}\} \cup V_{n+1})\}$. Hence, we have

$$\mathbb{P}\left\{V_{n+1} \le Q(\alpha; \hat{\mathcal{F}}) | \mathcal{X}\right\} = \mathbb{P}\left\{V_{n+1} \le Q(\frac{\lceil (n_1+n_0+1)\alpha \rceil - n_0}{n_1+1}; \{V_i : X_i = X_{n+1}\} \cup V_{n+1})\} \middle| \mathcal{X}\right\}$$
$$\overset{(b)}{=} \frac{\lceil (n_1+n_0+1)\alpha \rceil - n_0}{n_1+1} \le \frac{1}{n_1+1} + \frac{n_0+n_1+1}{n_1+1}(\alpha - \frac{n_0}{n_0+n_1+1}),$$

where step (b) uses the fact that $V_i \sim \text{Unif}[-1, 1]$ for all i with $X_i = X_{n+1}$. Hence, (B.1) holds. □

Next, we marginalize over $X_{1:n}$ but conditional on $m = n_0 + n_1$ (the total number of samples with $X_i \in \{0, X_{n+1}\}$). From (B.1):

$$\begin{aligned}\mathbb{P}\left\{V_{n+1} \le Q(\alpha; \hat{\mathcal{F}}) | m, X_{n+1}\right\} &\le \mathbb{E}[\frac{1}{n_1+1}|m] + \mathbb{E}\left[[\alpha\frac{m+1}{n_1+1} - \frac{n_0}{n_1+1}]_+ | m\right], \\ &= \mathbb{E}[\frac{1}{n_1+1}|m] + \mathbb{E}\left[[\alpha\frac{m+1-n_0}{n_1+1} - (1-\alpha)\frac{n_0}{n_1+1}]_+ | m\right], \\ &= \mathbb{E}[\frac{1}{n_1+1}|m] + (1-\alpha)\mathbb{E}\left[[\frac{\alpha}{1-\alpha} - \frac{n_0}{n_1+1}]_+ | m\right]. \qquad \text{(B.2)}\end{aligned}$$

Notice that conditional on m, X_i falls at 0 or X_{n+1} following an independent Bernoulli law:

$$X_i = \begin{cases} 0 & \text{w.p. } \frac{q_0}{q_0+q_1} = \alpha, \\ X_{n+1} & \text{w.p. } \frac{q_1}{q_0+q_1} = 1-\alpha. \end{cases}$$

From direct calculations, we obtain that

$$\begin{aligned}\mathbb{E}[\frac{1}{n_1+1}|m] &= \sum_{n_1=0}^{m} \frac{1}{n_1+1} \frac{m!}{n_1!(m-n_1)!}(1-\alpha)^{n_1}\alpha^{m-n_1} \\ &= \frac{1}{(m+1)(1-\alpha)} \sum_{n_1=1}^{m+1} \frac{(m+1)!}{n_1!(m+1-n_1)!}(1-\alpha)^{n_1}\alpha^{m+1-n_1} \le \frac{1}{m(1-\alpha)}. \qquad \text{(B.3)}\end{aligned}$$

Also, we have

$$\begin{aligned}&\mathbb{E}\left[[\frac{\alpha}{1-\alpha} - \frac{n_0}{n_1+1}]_+ | m\right] \\ \overset{(c)}{=} &\frac{\alpha}{1-\alpha}\mathbb{P}(n_0 \le \alpha(m+1)|m) - \sum_{n_0=1}^{n_0 \le \alpha(m+1)} \frac{n_0}{m-n_0+1} \frac{m!}{n_0!(m-n_0)!}\alpha^{n_0}(1-\alpha)^{m-n_0} \\ = &\frac{\alpha}{1-\alpha}\left(\mathbb{P}(n_0 \le \alpha(m+1)|m) - \sum_{n_0=0}^{n_0 \le \alpha(m+1)-1} \frac{m!}{n_0!(m-n_0)!}\alpha^{n_0}(1-\alpha)^{m-n_0}\right) \\ = &\frac{\alpha}{1-\alpha}\left(\mathbb{P}(n_0 \le \alpha(m+1)|m) - \mathbb{P}(n_0 \le \alpha(m+1) - 1|m)\right) \\ = &\frac{\alpha}{1-\alpha}\mathbb{P}(n_0 = \underbrace{\lfloor \alpha(m+1) \rfloor}_{n_*}|m) = \frac{\alpha}{1-\alpha}\binom{m}{n_*}\alpha^{n_*}(1-\alpha)^{m-n_*}. \qquad \text{(B.4)}\end{aligned}$$

We now use the Stirling's approximation:

$$\sqrt{2\pi k}(\frac{k}{e})^k e^{\frac{1}{12k+1}} \le k! \le \sqrt{2\pi k}(\frac{k}{e})^k e^{\frac{1}{12k}}, \quad \text{for all } k \ge 1.$$

Plug the Stirling's approximation into (B.4), there exist a constant $C > 0$ such that when $m \geq C$, we have:

$$\mathbb{E}\left[[\frac{\alpha}{1-\alpha} - \frac{n_0}{n_1+1}]_+|m\right] \leq \frac{\alpha}{1-\alpha}\exp(\frac{1}{12m})\sqrt{\frac{m}{2\pi(n^*)(m-n^*)}}(\alpha\frac{m}{n^*})^{n^*}\left((1-\alpha)\frac{m}{m-n^*}\right)^{m-n^*}$$
$$\overset{(c)}{\leq} \frac{\alpha}{1-\alpha}\exp(\frac{1}{12m})\sqrt{\frac{m}{2\pi(m\alpha-1)(m(1-\alpha)-1)}}(\frac{m\alpha}{m\alpha-1})^{n^*}\left(\frac{(1-\alpha)m}{m(1-\alpha)-1}\right)^{m-n^*}$$
$$\leq C\sqrt{\frac{1}{m}}(1+\frac{2}{m})^m \leq \frac{Ce^2}{\sqrt{m}}, \tag{B.5}$$

where we have used the fact that $m\alpha + 1 \leq n^* \geq \alpha m - 1$ at step (c). Notice that m itself follows a binomial distribution with n trials and successful rate $(q_1 + q_0)$. Apply the Chernoff bound, we have

$$\mathbb{P}\left\{m \leq \frac{(q_1+q_0)n}{2}\right\} \leq \exp(-\frac{n(q_1+q_0)}{8}). \tag{B.6}$$

For any constant $p \geq 1$, $n(q_1 + q_0) \to \infty$. Combine it with (B.2), (B.3), (B.5) and (B.6), there exist a constant $C > 0$, such that for all $X_{n+1} \in \{\pm e_j, j = 1, \dots, p\}$, we have

$$\mathbb{P}\left\{V_{n+1} \leq Q(\alpha;\hat{\mathcal{F}})|X_{n+1}\right\}$$
$$\leq \mathbb{P}\left\{\{V_{n+1} \leq Q(\alpha;\hat{\mathcal{F}})|X_{n+1}\} \cap \{m \geq \frac{(q_1+q_0)n}{2}\}\right\} + \mathbb{P}\left\{m \geq \frac{(q_1+q_0)n}{2}|X_{n+1}\right\} \leq C\sqrt{\frac{1}{(q_1+q_0)n}}.$$

Marginalize over X_{n+1}, we reach the desired result: there exists a sufficiently large constant C, such that

$$\mathbb{P}\left\{V_{n+1} \leq Q(\alpha;\hat{\mathcal{F}})\right\} \leq \mathbb{P}\left\{X_{n+1} \neq 0\right\}\frac{C}{\sqrt{(q_0+q_1)n}} + \mathbb{P}\left\{X_{n+1} = 0\right\} \leq \frac{C}{\sqrt{(q_0+q_1)n}} + q_0 \to q_0.$$

B.3. *Proof of Theorem 1*

Proof. Define

$$\mathcal{T} := \{\{Z_i, i = 1, \dots, n+1\} = \{z_i := (x_i, y_i), i = 1, \dots, n+1\}\}.$$

Let σ be a permutation of numbers $1, 2, \dots, n+1$ that specifies how the values are assigned, e.g., Z_i takes value z_{σ_i}. Since $V(.;\mathcal{Z})$ and $H(.,.;\mathcal{X})$ are fixed conditional on $\mathcal{T}$, we can set $v^*_{\sigma_i} = Q(\tilde{\alpha}; \sum_{j=1}^n p^H_{\sigma_i j}\delta_{v_j})$ as the realized empirical quantile at $\tilde{\alpha}$ for $\hat{\mathcal{F}}_i$ given a particular permutation ordering σ. Hence, for any given $\tilde{\alpha} \in \Gamma$, conditional $\mathcal{T}$ and the permutation ordering σ, we have

$$\sum_{i=1}^{n+1}\mathbb{1}_{V_i \leq Q(\tilde{\alpha};\hat{\mathcal{F}}_i)}|\mathcal{T},\sigma = \sum_{i=1}^{n+1}\mathbb{1}_{v_{\sigma_i} \leq v^*_{\sigma_i}} = \sum_{i=1}^{n+1}\mathbb{1}_{v_i \leq v^*_i}. \tag{B.7}$$

In other words, the achieved value for the left side of (B.7) or Theorem 1 (4) remains the same for all σ. Since Γ is fixed conditional on $\mathcal{T}$, the smallest value in Γ satisfying (4) is also fixed conditional on $\mathcal{T}$, by Lemma A.3, we obtain that

$$\mathbb{P}\{V_{n+1} \leq Q(\tilde{\alpha};\hat{\mathcal{F}}_{n+1})|\mathcal{T}\} = \mathbb{E}\left[\frac{1}{n+1}\sum_{i=1}^{n+1}\mathbb{1}_{v_i \leq v^*_i}|\mathcal{T}\right] = \frac{1}{n+1}\sum_{i=1}^{n+1}\mathbb{1}_{v_i \leq v^*_i} \geq \alpha$$

Marginalize over $\mathcal{T}$, we have

$$\mathbb{P}\{V_{n+1} \leq Q(\tilde{\alpha};\hat{\mathcal{F}}_{n+1})\} \geq \alpha. \tag{B.8}$$

By Lemma A.1, equivalently, we also have

$$\mathbb{P}\{V_{n+1} \leq Q(\tilde{\alpha};\hat{\mathcal{F}})\} \geq \alpha. \tag{B.9}$$

B.4. *Proof of Theorem 2*

Define

$$\mathcal{T} \coloneqq \{\{Z_i, i = 1, \ldots, n+1\} = \{z_i \coloneqq (x_i, y_i), i = 1, \ldots, n+1\}\}.$$

By (B.7) and the fact that Γ is fixed conditional on $\mathcal{T}$, we know that $\tilde{\alpha}_1$, $\tilde{\alpha}_2$ and α_1, α_2 are fixed conditional on $\mathcal{T}$. As a result, when $\tilde{\alpha} = \begin{cases} \tilde{\alpha}_1 \ w.p. \ \frac{\alpha - \alpha_2}{\alpha_1 - \alpha_2} \\ \tilde{\alpha}_2 \ w.p. \ \frac{\alpha_1 - \alpha}{\alpha_1 - \alpha_2} \end{cases}$, and it is independent of the data conditional on $\mathcal{T}$. Apply Lemma A.3, we have

$$\begin{aligned} \mathbb{P}\{V_{n+1} \le Q(\tilde{\alpha}; \hat{\mathcal{F}}_{n+1}) | \mathcal{T}\} &= \mathbb{E}\left[\frac{1}{n+1} \sum_{i=1}^{n+1} \mathbb{1}_{v_i \le v_i^*} | \mathcal{T}\right] \\ &= \alpha_1 \frac{\alpha - \alpha_2}{\alpha_1 - \alpha_2} + \alpha_2 \frac{\alpha_1 - \alpha}{\alpha_1 - \alpha_2} = \alpha. \end{aligned}$$

Marginalizing over $\mathcal{T}$, we have

$$\mathbb{P}\{V_{n+1} \le Q(\tilde{\alpha}; \hat{\mathcal{F}}_{n+1})\} = \alpha.$$

By Lemma A.1, equivalently, we have

$$\mathbb{P}\{V_{n+1} \le Q(\tilde{\alpha}; \hat{\mathcal{F}})\} = \alpha.$$

B.5. *Proof of Lemma 1*

Proof. As a direct application of Theorem 1 and Crorllary 1, we obtain that

$$\mathbb{P}\left\{V_{n+1} \in C_V(X_{n+1})\right\} \ge \alpha, \ \mathbb{P}\left\{Y_{n+1} \in C(X_{n+1})\right\} \ge \alpha.$$

The fact that $C_V(X_{n+1})$ is an interval comes directly from Lemma A.2 (iii). □

B.6. *Proof of Lemma 2*

Proof.

– Proof of part 1: By definition, $V_{n+1} = v \in C_V(X_{n+1})$ iff (if and only if) the smallest value $\tilde{\alpha} \in \Gamma$ that makes (6) hold is greater than $\sum_{V_i < v} p^H_{n+1,i} \in \Gamma$. That is, $v \in C_V(X_{n+1})$ iff

$$\frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}_{V_i \le Q(\sum_{V_i < v} p^H_{n+1,i}; \hat{\mathcal{F}}_i(v))} < \alpha. \tag{B.10}$$

(a) When $v = \overline{V}_k$ for some $1 \le k \le n+1$, $\sum_{V_i < \overline{V}_k} p^H_{n+1,i} = \tilde{\theta}_k$ by definition. Hence $v \in C_V(X_{n+1})$ iff

$$\frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}_{V_i \le Q(\tilde{\theta}_k; \hat{\mathcal{F}}_i(\overline{V}_k))} < \alpha. \tag{B.11}$$

(b) When $v \in (\overline{V}_{\ell(k)}, \overline{V}_k)$ for some $1 \le k \le n+1$, $\sum_{V_i < v} p^H_{n+1,i} = \sum_{V_i < \overline{V}_k} p^H_{n+1,i} = \tilde{\theta}_k$. Hence $v \in C_V(X_{n+1})$ iff

$$\frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}_{V_i \le Q(\tilde{\theta}_k; \hat{\mathcal{F}}_i(v))} < \alpha. \tag{B.12}$$

A key observation is that the status of event $\{V_i \le Q(\tilde{\theta}_k; \hat{\mathcal{F}}_i(v))\}$ does not change as we vary $v \in [\overline{V}_{\ell(k)}, \overline{V}_k)$. That is, for all $1 \le i \le n$, we have

$$J_{ik}(v) \coloneqq \{V_i \le Q(\tilde{\theta}_k; \hat{\mathcal{F}}_i(v))\} = \{V_i \le Q(\tilde{\theta}_k; \hat{\mathcal{F}}_i(\overline{V}_{\ell(k)}))\} \coloneqq J_{ik}.$$

This can be easily verified:

◇ If $V_i < \overline{V}_k$, we have $V_i \leq \overline{V}_{\ell(k)} < v$, and

$$\{V_i \leq Q(\tilde{\theta}_k; \hat{\mathcal{F}}_i(v))\} = \{\tilde{\theta}_k > \theta_i\} = \{V_i \leq Q(\tilde{\theta}_k; \hat{\mathcal{F}}_i(\overline{V}_{\ell(k)}))\}$$

◇ If $V_i \geq \overline{V}_k$, then $V_i > v > \overline{V}_{\ell(k)}$, and we have

$$\{V_i \leq Q(\tilde{\theta}_k; \hat{\mathcal{F}}_i(v))\} = \{\tilde{\theta}_k > \theta_i + p^H_{i,n+1}\} = \{V_i \leq Q(\tilde{\theta}_k; \hat{\mathcal{F}}_i(\overline{V}_{\ell(k)}))\},$$

Hence, we obtain

$$\frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}_{V_i \leq Q(\tilde{\theta}_k; \hat{\mathcal{F}}_i(\overline{V}_{l(k)}))} < \alpha. \tag{B.13}$$

Combine part (a) and part (b), and the fact that $\overline{V}_{\ell(k)} \leq \overline{V}_k$ and $Q(\tilde{\theta}_k; \hat{\mathcal{F}}_i(v))$ is non-decreasing in v (Lemma A.2), we immediately reach the desired result that

$$\bar{C}^V(X_{n+1}) = \{v : v \leq Q(\tilde{\theta}_{k^*}; \hat{\mathcal{F}})\},$$

where k^* is the largest value of k such that (B.13) holds.

– Proof of part 2: As we increase k, both $\overline{V}_{l(k)}$ and $\tilde{\theta}_k$ are non-decreasing, hence, $Q(\tilde{\theta}_k; \hat{\mathcal{F}}_i(\overline{V}_{\ell(k)}))$ is non-decreasing in k. Thus, $J_{ik} = \{V_i \leq Q(\tilde{\theta}_k; \hat{\mathcal{F}}_i(\bar{V}_{\ell(k)}))\}$ is a monotone event in k: for all $k' \geq k$, we have $J_{ik} \subseteq J_{ik'}$. Consequently, suppose k_i^* is when J_{ik} first holds, then $\mathbb{1}_{J_{ik}} = 1$ iff $k \geq k_i^*$. We can divide J_{ik} into two subsets:

$$\begin{aligned} J_{ik} &= \Big(\{V_i > \bar{V}_{\ell(k)}\} \cap \{V_i \leq Q(\tilde{\theta}_k; \hat{\mathcal{F}}_i(\bar{V}_{\ell(k)}))\}\Big) \cup \Big(\{V_i \leq \bar{V}_{\ell(k)}\} \cap \{V_i \leq Q(\tilde{\theta}_k; \hat{\mathcal{F}}_i(\bar{V}_{\ell(k)}))\}\Big) \\ &\overset{(a)}{=} \underbrace{\Big(\{\ell(i) \geq \ell(k)\} \cap \{\theta_i + p^H_{i,n+1} < \tilde{\theta}_k\}\Big)}_{J^1_{ik}} \cup \underbrace{\Big(\{\ell(i) < \ell(k)\} \cap \{\theta_i < \tilde{\theta}_k\}\Big)}_{J^2_{ik}}. \end{aligned} \tag{B.14}$$

At step (a), we have used the fact that

$$V_i > \overline{V}_{\ell(k)} \Leftrightarrow \ell(i) \geq \ell(k),$$

and that

– when $V_i > \overline{V}_{\ell(k)}$, we have $\sum_{1 \leq j \leq n+1: V_j < V_i} p^H_{ij} = \theta_i + p^H_{i,n+1}$ when $V_{n+1} = \bar{V}_{\ell(k)}$ Hence,

$$V_i \leq Q(\tilde{\theta}_k; \hat{\mathcal{F}}_i(\bar{V}_{l(k)})) \Leftrightarrow \theta_i + p^H_{i,n+1} < \tilde{\theta}_k.$$

– when $V_i \leq \overline{V}_{\ell(k)}$, we have $\sum_{1 \leq j \leq n+1: V_j < V_i} p^H_{ij} = \theta_i$ when $V_{n+1} = \bar{V}_{\ell(k)}$. Hence,

$$V_i \leq Q(\tilde{\theta}_k; \hat{\mathcal{F}}_i(\bar{V}_{l(k)})) \Leftrightarrow \theta_i < \tilde{\theta}_k.$$

We now consider when J_{ik} turns true for samples from categories A_1, A_2 and A_3.

– For $i \in A_1$, by the definition of A_1 and (B.14), we know that J_{ik} is true at $k = i$ and $k_i^* \leq i$, and $J_{ik} = J^1_{ik} = \{\theta_i + p^H_{i,n+1} < \tilde{\theta}_k\}$.

– For $i \in A_2 \cup A_3$, since $i \notin A_1$ and $k_i^* > i$, J^1_{ik} fails to hold for all k. Hence, J_{ik} holds when J^2_{ik} holds.

◇ When $i \in A_2$: since $\theta_i \geq \tilde{\theta}_i$, in order for $\theta_i < \tilde{\theta}_k$ to hold, by definition, we must have $\sum_{j \leq l(k)} p^H_{n+1,j} = \tilde{\theta}_k > \tilde{\theta}_i = \sum_{j \leq l(i)} p^H_{n+1,j}$, which automatically guarantees that $l(k) > l(i)$. As a result, for $i \in A_2$, we have $J_{ik} = J^2_{ik} = \{\theta_i < \tilde{\theta}_k\}$.

◇ When $i \in A_3$: in order to have $l(k) > l(i)$, we automatically $\tilde{\theta}_k \geq \tilde{\theta}_i > \theta_i$ for samples in A_3. Thus, for $i \in A_3$, we have $J_{ik} = J^2_{ik} = \{l(i) < l(k)\}$.

Combine them together, we have

$$
\begin{aligned}
S(k) &= \sum_{i=1}^{n} \frac{1}{n+1} \mathbb{1}_{V_i \le Q(\tilde{\theta}_k; \hat{\mathcal{F}}_i(\overline{V}_{\ell(k)}))} \\
&= \frac{1}{n+1}\left(\sum_{i\in A_1} \mathbb{1}_{J^1_{ik}} + \sum_{i\in A_2} \mathbb{1}_{J^2_{ik}} + \sum_{i\in A_3} \mathbb{1}_{J^2_{ik}} \right) \\
&= \frac{1}{n+1}\left(\sum_{i\in A_1} \mathbb{1}_{\{\theta_i + p^H_{i,n+1} < \tilde{\theta}_k\}} + \sum_{i\in A_2} \mathbb{1}_{\theta_i < \tilde{\theta}_k} + \sum_{i\in A_3} \mathbb{1}_{l(i) < l(k)} \right).
\end{aligned}
$$

We have proved the second part of Lemma 2. □

LOCAL COVERAGE PROPERTIES OF LCP

B.7. Proof of Theorem 3

Proof. We first prove the convergence from $\tilde{\alpha}(v)$ to α in (11) and then show that the achieved coverage levels converge to the nominal level for both $\tilde{\alpha} = \alpha$ and $\tilde{\alpha} = \tilde{\alpha}(v)$ as described in Lemma 1. Define $I_i = \frac{\sum_{j=1, j\neq i}^{n} P_{V|X_j}(V_i) H_{ij}}{B_i}$ and $R_i = \sum_{j=1,j\neq i}^{n} \frac{H_{ij}(\mathbb{1}_{V_j < V_i} - P_{V|X_j}(V_i))}{B_i}$ for all $i = 1, \dots, n+1$.

1. **Proof of (11):** For $i = 1, \dots, n$, define $B_i = \sum_{j=1}^{n+1} H_{ij}$ and, for any $\tilde{\alpha} \in [0,1]$ and $v \in \mathbb{R}$, define

$$J_i(v, \tilde{\alpha}) \coloneqq \{V_i \le Q(\tilde{\alpha}; \hat{\mathcal{F}}(v))\} = \{\tilde{\alpha} > \frac{\sum_{j\le n: V_j < V_i} H_{ij} + \mathbb{1}_{v < V_i}}{B_i}\}.$$

$J_i(v, \tilde{\alpha})$ is the event for wether sample i contributes to the left side of Lemma 1 (6). We can define a subset event $\underline{J}_i(\tilde{\alpha}) \subseteq J_i(v, \tilde{\alpha})$ for all v values for all v. Decompose the condition of $J_i(v, \tilde{\alpha})$ as below:

$$\frac{\sum_{j\le n: V_j < V_i} H_{ij} + \mathbb{1}_{v < V_i}}{B_i} \le \frac{\sum_{j\le n: V_j < V_i} H_{ij}}{B_i} + \frac{1}{B_i} = I_i + R_i + \frac{1}{B_i}. \tag{B.15}$$

Set $G = \left\{ i \in \{1, \dots, n\} : B_i \ge \frac{1}{2eL} n h_n^{\beta}, |R_i| \le \sqrt{\frac{2eL \ln n}{n h_n^{\beta}}} \right\}$. By Lemma A.4 (i), there exists a constant $C > 0$, such that for all $i \in G$:

$$\frac{B_i - 1 - H_{i,n+1}}{B_i} \in [1 - \frac{4eL}{n h_n^{\beta}}, 1], \quad |I_i - \frac{B_i - 1 - H_{i,n+1}}{B_i} P_{V|X_i}(V_i)| \le C h_n \ln(h_n^{-1}). \tag{B.16}$$

Combine (B.15) with (B.16), there exist a constant $C > 0$ such that for all $i \in G$, we have

$$\underline{J}_i(\tilde{\alpha}) \coloneqq \left\{ \tilde{\alpha} > P_{V|X_i}(V_i) + C\left(h_n \ln(h_n^{-1}) + \sqrt{\frac{\ln n}{n h_n^{\beta}}} \right) \right\} \subseteq J_i(v, \tilde{\alpha}), \text{ for all } v. \tag{B.17}$$

We can also define a superset event $\bar{J}_i(\tilde{\alpha}) \supseteq J_i(v, \tilde{\alpha})$ for all v values:

$$\frac{\sum_{j\le n: V_j < V_i} H_{ij} + \mathbb{1}_{v < V_i}}{B_i} \ge \frac{\sum_{j\le n: V_j < V_i} H_{ij}}{B_i} = I_i + R_i. \tag{B.18}$$

Combine (B.18) with (B.16), there exists a constant $C > 0$ such that for all $i \in G$, we have

$$J_i(v, \tilde{\alpha}) \subseteq \bar{J}_i(\tilde{\alpha}) \coloneqq \left\{ \tilde{\alpha} > P_{V|X_i}(V_i) - C\left(h_n \ln(h_n^{-1}) + \sqrt{\frac{\ln n}{n h_n^{\beta}}} \right) \right\}, \text{ for all } v. \tag{B.19}$$

Hence, we can then upper and lower bound the left side of (6) using $\bar{J}_i(\tilde{\alpha})$ and $\underline{J}_i(\tilde{\alpha})$:

$$\frac{1}{n+1}\sum_{i=1}^{n+1} J_i(v,\tilde{\alpha}) \le \frac{1}{n+1} + \frac{1}{n+1}\sum_{i\in G}\bar{J}_i(\tilde{\alpha}) + \frac{|G^c|}{n+1}, \tag{B.20}$$

$$\frac{1}{n+1}\sum_{i=1}^{n+1} J_i(v,\tilde{\alpha}) \ge \frac{1}{n+1}\sum_{i\in G}\underline{J}_i(\tilde{\alpha}). \tag{B.21}$$

Set $W_i = P_{V|X_i}(V_i)$, which is i.i.d generated from $\mathrm{Unif}[0,1]$ when $V|X_i$ is a continuous variable. By Lemma A.4, we know that

$$\begin{aligned}
\mathbb{P}\{|G^c|>0\} &\le \mathbb{P}\{\min_{i=1}^{n} B_i \le \frac{nh_n^\beta}{2eL}\} + \mathbb{P}\{\exists i\in\{1,\dots,n\}: B_i > \frac{nh_n^\beta}{2eL}, |R_i| > \sqrt{\frac{2eL\ln n}{nh_n^\beta}}\} \\
&\le \mathbb{P}\{\min_{i=1}^{n} B_i \le \frac{nh_n^\beta}{2eL}\} + n\max_{1\le i\le n}\mathbb{P}\{|R_i| > \sqrt{\frac{\ln n}{B_i}}\} \\
&\le n\exp(-\frac{nh_n^\beta}{8L}) + n\times\frac{1}{n^2} \to 0.
\end{aligned} \tag{B.22}$$

When $\{|G^c|=0\}$ holds:

– When $\tilde{\alpha}$ makes (6) hold, by (B.20), we must have

$$\frac{1}{n+1}\left(1+\sum_{i=1}^{n}\bar{J}_i(\tilde{\alpha})\right) \ge \alpha \Rightarrow \tilde{\alpha} \ge Q(\frac{n+1}{n}\alpha - \frac{1}{n}; \frac{1}{n}\sum_{i=1}^{n}\delta_{W_i}) - C\left(h_n\ln(h_n^{-1}) + \sqrt{\frac{\ln n}{nh_n^\beta}}\right). \tag{B.23}$$

– By (B.21), $\tilde{\alpha}$ makes (6) hold as long as

$$\frac{1}{n+1}\sum_{i=1}^{n}\underline{J}_i(\tilde{\alpha}) \ge \alpha \Rightarrow \tilde{\alpha} \ge Q(\frac{n+1}{n}\alpha; \frac{1}{n}\sum_{i=1}^{n}\delta_{W_i}) + C\left(h_n\ln(h_n^{-1}) + \sqrt{\frac{\ln n}{nh_n^\beta}}\right).$$

Further, since Γ includes all possible empirical CDF values from weighted distribution $\hat{\mathcal{F}}_i$ for $i=1,\dots,n+1$ under all possible ordering of $V_1,\dots,V_{n+1}$. Let $B_{\max} = \max_{i=1}^{n+1} B_i$. The differences between two adjacent values in Γ is upper bounded by $\frac{1}{B_{\max}} \le \frac{2eL}{nh_n^\beta}$. Hence, there exists a constant $C>0$ such that the smallest value in Γ that makes (6) is upper bounded by

$$\tilde{\alpha} \le Q(\frac{n+1}{n}\alpha; \frac{1}{n}\sum_{i=1}^{n}\delta_{W_i}) + C\left(h_n\ln(h_n^{-1}) + \sqrt{\frac{\ln n}{nh_n^\beta}}\right). \tag{B.24}$$

The bounds (B.23) and (B.24) hold for all $V_{n+1}=v$. By Dvoretzky–Kiefer–Wolfowitz inequality and the fact that $W_i \sim \mathrm{Unif}[0,1]$, there exists a constant $C>0$ such that

$$\mathbb{P}(\max_t |Q(t;\sum_{i=1}^{n}\delta_{W_i}) - t| \le \sqrt{\frac{\ln n}{n}}) \le \frac{C}{n^2}. \tag{B.25}$$

Combine (B.23), (B.24), (B.25) and (B.22), there exist a constant $C>0$, such that

$$\mathbb{P}\left\{|\min_{v_{n+1}}\tilde{\alpha}(v_{n+1}) - \alpha| < C\left(h_n\ln(h_n^{-1}) + \sqrt{\frac{\ln n}{nh_n^\beta}}\right)\right\} \ge \frac{C}{n^2} + \mathbb{P}(|G^c|>0) \to 0. \tag{B.26}$$

Since $C\left(h_n\ln(h_n^{-1}) + \sqrt{\frac{\ln n}{nh_n^\beta}}\right) \to 0$, this concludes our proof.

2. **Proofs of (9) and (10):** By definition, for any given $\tilde{\alpha}$, $V_{n+1} \leq Q(\tilde{\alpha}; \hat{\mathcal{F}})$ if and only if

$$\sum_{i=1}^{n} \frac{H(X_{n+1}, X_i)}{B_{n+1}} \mathbb{1}_{V_i < V_{n+1}} = I_{n+1} + R_{n+1} < \tilde{\alpha}. \tag{B.27}$$

Define $G = \{B_{n+1} \geq \frac{nh_n^\beta}{2eL}, |R_{n+1}| \leq \sqrt{\frac{2eL \ln n}{nh_n^\beta}}\}$. When G holds, following the same routine as bounding $J(\tilde{\alpha}, v)$ with $\bar{J}(\tilde{\alpha})$ and $\underline{J}(\tilde{\alpha})$, we can lower and upper bound $(I_{n+1} + R_{n+1})$ in (B.27) using Lemma A.4 (i): there exists a constant $C > 0$, such that

$$I_{n+1} + R_{n+1} \leq P_{V|X_{n+1}}(V_{n+1}) + C\left(h_n \ln(h_n^{-1}) + \sqrt{\frac{\ln n}{nh_n^\beta}}\right). \tag{B.28}$$

$$I_{n+1} + R_{n+1} \geq P_{V|X_{n+1}}(V_{n+1}) - C\left(h_n \ln(h_n^{-1}) + \sqrt{\frac{\ln n}{nh_n^\beta}}\right). \tag{B.29}$$

$W_{n+1} = P_{V|X_{n+1}}(V_{n+1}) \sim \text{Unif}[0,1]$ since $V|X_{n+1}$ is a continuous variable. By Lemma A.4 (i) and (ii), $\mathbb{P}(G^c) \to 0$. Hence, for any given $\tilde{\alpha}$, there exists a constant $C > 0$, such that

$$\mathbb{P}(I_{n+1} + R_{n+1} < \tilde{\alpha}) \leq \tilde{\alpha} + C\left(h_n \ln(h_n^{-1}) + \frac{1}{(nh_n^\beta)^{1/3}}\right) + \mathbb{P}\left\{G^c\right\} \to \tilde{\alpha}, \tag{B.30}$$

$$\mathbb{P}(I_{n+1} + R_{n+1} < \tilde{\alpha}) \geq \tilde{\alpha} - C\left(h_n \ln(h_n^{-1}) + \frac{1}{(nh_n^\beta)^{1/3}}\right) \to \tilde{\alpha}. \tag{B.31}$$

Consequently, when $\tilde{\alpha} = \alpha$ or $\tilde{\alpha} = \tilde{\alpha}(v) \to \alpha$ for all v in probability as described in (11), we achieve an asymptotic conditional coverage at level α. □

B.8. Proof of Theorem 4

Proof. We use the result from Barber et al. (2019) which extends CP to the setting with covariate shift:

Proposition B.1 (Barber et al. (2019), Corollary 1). For any fixed x_0. Set $w_{x_0}(.) = \frac{d\tilde{\mathcal{P}}_X^{x_0}}{d\mathcal{P}_X}$ and $p_i^{x_0}(x) = \frac{w_{x_0}(X_i)}{\sum_{j=1}^n w_{x_0}(X_i) + w_{x_0}(x)}$ for $i = 1, \ldots, n$, and $p_{n+1}^{x_0}(x) = \frac{w_{x_0}(x)}{\sum_{j=1}^n w_{x_0}(X_i) + w_{x_0}(x)}$. Then,

$$\mathbb{P}\left\{V(X_{n+1}, Y_{n+1}) \leq Q(\alpha; \sum_{i=1}^{n+1} p_i^{x_0}(X_{n+1})\delta_{\overline{V}_i})\right\} \geq \alpha.$$

In our setting, $w_{x_0}(x) \propto H(x_0, x)$. As a direct application of Proposition B.1, when $(\tilde{X}, \tilde{Y})$ is distributed from $\tilde{\mathcal{P}}_{XY}^{X_{n+1}}$, we have

$$\mathbb{P}\left\{V(\tilde{X}, \tilde{Y}) \leq Q(\alpha; \sum_{i=1}^{n+1} p_i^{x_0}(\tilde{X})\delta_{\overline{V}_i})|X_{n+1} = x_0\right\} \geq \alpha.$$

Since the $H(x_0, x_0) \geq H(x_0, \tilde{X})$ by definition, the distribution $\hat{\mathcal{F}}$ dominates the distribution $\sum_{i=1}^{n+1} p_i^{x_0}(\tilde{X})\delta_{\overline{V}_i}$: given $X_{n+1} = x_0$, for any α, we have

$$\begin{aligned}
Q(\alpha; \hat{\mathcal{F}}) &= Q(\alpha; \sum_{i=1}^{n+1} \frac{H(x_0, X_i)}{\sum_{j=1}^{n+1} H(x_0, X_j)} \delta_{\bar{V}_i}) \\
&\geq Q(\alpha; \sum_{i=1}^{n} \frac{H(x_0, X_i)}{\sum_{j=1}^{n} H(x_0, X_j) + H(x_0, \tilde{X})} \delta_{\bar{V}_i} + \frac{H(x_0, \tilde{X})}{\sum_{j=1}^{n} H(x_0, X_j) + H(x_0, \tilde{X})} \delta_{\bar{V}_i}) = Q(\alpha; \sum_{i=1}^{n+1} p_i^{x_0}(\tilde{X})\delta_{\overline{V}_i}).
\end{aligned}$$

Hence, we have

$$\mathbb{P}\{V(\tilde{X},\tilde{Y}) \le Q(\alpha;\hat{\mathcal{F}})|X_{n+1}=x_0\} \ge \alpha, \text{ for all } x_0.$$

Next, we turn to the achieved coverage using $\tilde{C}(X_{n+1})$. By construction, we have

$$\begin{aligned}\{\tilde{Y} \in \tilde{C}(X_{n+1})\} &= \{V(X_{n+1},\tilde{Y}) \le Q(\alpha;\hat{\mathcal{F}}) + \varepsilon(X_{n+1})\}\\ &\supseteq \{V(\tilde{X},\tilde{Y}) \le Q(\alpha;\hat{\mathcal{F}})\}.\end{aligned}$$

Consequently, we obtain

$$\mathbb{P}\left\{\tilde{Y} \in \tilde{C}(X_{n+1})|X_{n+1}=x_0\right\} \ge \mathbb{P}\left\{V(\tilde{X}) \le Q(\alpha;\hat{\mathcal{F}}))|X_{n+1}=x_0\right\} \ge \alpha.$$

C. PROOF OF LEMMAS IN THE APPENDIX

C.1. *Proof of Lemma A.1*

Proof. By definition, we know

$$V_{n+1} \le Q(\alpha;\sum_{i=1}^{n} w_i\delta_{V_i} + w_{n+1}\delta_{V_{n+1}}) \Rightarrow V_{n+1} \le Q(\alpha;\sum_{i=1}^{n} w_i\delta_{V_i} + w_{n+1}\delta_{\infty}).$$

To show that Lemma A.1 holds, we only need to show that,

$$V_{n+1} > Q(\alpha;\sum_{i=1}^{n} w_i\delta_{V_i} + w_{n+1}\delta_{V_{n+1}}) \Rightarrow V_{n+1} > Q(\alpha;\sum_{i=1}^{n} w_i\delta_{V_i} + w_{n+1}\delta_{\infty}).$$

Let $Q(\alpha;\sum_{i=1}^{n} w_i\delta_{V_i} + w_{n+1}V_{n+1}) = V_{i^*}$ for some index $1 \le i^* \le n+1$. When $V_{n+1} > V_{i^*}$, we must have $V_{i^*} < \infty$. By definition:

$$\begin{aligned}\alpha \ge \sum_{i=1}^{n+1} w_i\mathbb{1}_{V_i \le V_{i^*}} &= \sum_{i=1}^{n} w_i\mathbb{1}_{V_i \le V_{i^*}} = \sum_{i=1}^{n} w_i\mathbb{1}_{V_i \le V_{i^*}} + w_{n+1}\mathbb{1}_{\infty \le V_{i^*}}\\ \Rightarrow Q(\alpha;\sum_{i=1}^{n} w_i\mathbb{1}_{V_i} + w_{n+1}\delta_{\infty}) &\le V_{i^*} < V_{n+1}.\end{aligned}$$

C.2. *Proof of Lemma A.2*

Proof. We can prove Lemma A.2 with elementary calculus arguments.

(i) Given $V_1,\dots,V_{n+1}$, $Q(\tilde{\alpha};\hat{\mathcal{F}}_i) = \inf\{t : \mathbb{P}(v \le t) \ge \tilde{\alpha}, v \sim \hat{\mathcal{F}}_i\}$. The empirical distribution $\hat{\mathcal{F}}$ is discrete with mass p_{ij}^H on V_i, we can have an explicit expression for $Q(\tilde{\alpha};\hat{\mathcal{F}}_i)$:

$$Q(\tilde{\alpha};\hat{\mathcal{F}}_i) = \begin{cases}\overline{V}_0, & \tilde{\alpha}=0,\\ \overline{V}_i, & \sum_{j=1}^{i-1} p_{ij}^H < \tilde{\alpha} \le \sum_{j=1}^{i} p_{ij}^H, i=1,\dots,n,\\ \overline{V}_{n+1}, & \sum_{j=1}^{n} p_{ij}^H < \tilde{\alpha}.\end{cases}$$

Hence, $Q(\tilde{\alpha};\hat{\mathcal{F}}_i)$ is non-decreasing and right-continuous piece-wise constant on $\tilde{\alpha}$, and v_i^* can only change its value at $\sum_{j=1}^{k} p_{ij}^H$ for $k=1,\dots,n$. The same is true for $Q(\tilde{\alpha};\hat{\mathcal{F}})$.

(ii) Given $\tilde{\alpha}$, when increasing V_{n+1} from $V_{n+1}=v'$ to $V_{n+1}=v$ for $v>v'$, the empirical distribution $\hat{\mathcal{F}}_i(v)$ dominates the empirical distribution $\hat{\mathcal{F}}_i(v')$ by construction: $\forall\tilde{\alpha}$, we have

$$\mathbb{P}\left\{t \le \tilde{\alpha}|t \sim \hat{\mathcal{F}}_i(v')\right\} \ge \mathbb{P}\left\{t \le \tilde{\alpha}|t \sim \hat{\mathcal{F}}_i(v)\right\}.$$

As a result, $Q(\tilde{\alpha};\hat{\mathcal{F}}_i(v))$ is non-decreasing on v for any given $\tilde{\alpha}$, for $i=1,\dots,n+1$.

(iii) Suppose that $v \in C_V(X_{n+1})$. Let $\tilde{\alpha} \in \Gamma$ be the smallest value such that

$$\sum_{i=1}^{n+1} \frac{1}{n+1} \mathbb{1}_{V_i \leq Q(\tilde{\alpha}; \hat{\mathcal{F}}_i(v))} \geq \alpha,$$

by definition, we have $v \leq Q(\tilde{\alpha}; \hat{\mathcal{F}})$. Now, we consider $V_{n+1} = v'$ for $v' \leq v$. By the monotonicity of $Q(\tilde{\alpha}; \hat{\mathcal{F}}_i(v))$ on $\tilde{\alpha}$ and v from Lemma A.2 (i) and (ii), we must have $\tilde{\alpha}' \geq \tilde{\alpha}^*$ where $\tilde{\alpha}' \in \Gamma$ is the smallest value satisfying

$$\sum_{i=1}^{n+1} \frac{1}{n+1} \mathbb{1}_{V_i \leq Q(\tilde{\alpha}'; \hat{\mathcal{F}}_i(v'))} \geq \alpha,$$

Hence, we have $v' \leq v \leq Q(\tilde{\alpha}; \hat{\mathcal{F}}) \leq Q(\tilde{\alpha}'; \hat{\mathcal{F}})$ and v' is included in the PI. This concludes our proof. □

C.3. Proof of Lemma A.3

Proof. Let σ be a permutation of numbers $1, 2, \ldots, n+1$. We know that

$$\mathbb{P}\left\{\sigma_{n+1} = i | \mathcal{T}\right\} = \frac{\#\{\sigma : \sigma_{n+1} = i\}}{\sum_{j=1}^{n+1} \#\{\sigma : \sigma_{n+1} = j\}} = \frac{1}{n+1}.$$

Set $\mathcal{X} = \{X_1, \ldots, X_{n+1}\}$ be the unordered set of the features. Since the function $V(., \mathcal{Z})$ and the localizer $H(., ., \mathcal{X})$ are fixed functions conditional on $\mathcal{T}$, and $\tilde{\alpha}$ (can be random) is independent of the data conditional $\mathcal{T}$, we obtain

$$\begin{aligned} &\mathbb{P}\left\{V_{n+1} \leq Q(\tilde{\alpha}; \sum_{j=1}^{n+1} p^H_{n+1,j} \delta_{V_j}) | \mathcal{T}, \tilde{\alpha}\right\} \\ =& \sum_{i=1}^{n+1} \mathbb{P}\{\sigma_{n+1} = i | \mathcal{T}\} \mathbb{1}_{\{V_{n+1} \leq v^*_{n+1}(\sigma) | \mathcal{T}, \sigma_{n+1} = i\}} \\ =& \sum_{i=1}^{n+1} \frac{1}{n+1} \mathbb{1}_{\{v_i \leq v^*_{n+1}(\sigma), \sigma_{n+1} = i\}} \end{aligned} \tag{C.1}$$

where

$$v^*_i(\sigma) \coloneqq Q(\tilde{\alpha}; \sum_{j=1}^{n+1} p^H_{\sigma_i, \sigma_j} \delta_{v_{\sigma_j}}) = Q(\tilde{\alpha}; \sum_{j=1}^{n+1} \frac{H(x_{\sigma_i}, x_{\sigma_j})}{\sum_{j'=1}^{n+1} H(x_{\sigma_i}, x_{\sigma_{j'}})} \delta_{v_{\sigma_j}})$$

is the realization of $v^*_i \coloneqq Q(\tilde{\alpha}; \hat{\mathcal{F}}_i)$ under permutation σ, conditional on $\mathcal{T}$ and $\tilde{\alpha}$. We immediately observe that,

$$v^*_i(\sigma) = v^*_{\sigma_i} \tag{C.2}$$

Combine (C.1) and (C.2), we obtain that $\mathbb{P}\{V_{n+1} \leq v^*_{n+1} | \mathcal{T}, \tilde{\alpha}\} = \sum_{i=1}^{n+1} \frac{1}{n+1} \mathbb{1}_{\{v_i \leq Q(\tilde{\alpha}; \hat{\mathcal{F}}_i)\}}$. Marginalize over $\tilde{\alpha} | \mathcal{T}$, we have

$$\mathbb{P}\left\{V_{n+1} \leq Q(\tilde{\alpha}; \hat{\mathcal{F}}_{n+1}) | \mathcal{T}\right\} = \mathbb{E}\left[\sum_{i=1}^{n+1} \frac{1}{n+1} \mathbb{1}_{\{v_i \leq Q(\tilde{\alpha}; \hat{\mathcal{F}}_i)\}} | \mathcal{T}\right].$$

C.4. Proof of Lemma A.4

Proof.

Part (i): We divide the space into non-overlapping subregions $A_k = \{x : (k-1)h_n \le d(x_0, x) < kh_n\}$. Then,

$$B(x_0) = \sum_{i=1}^{n+1} \exp(-\frac{d(x_0, X_i)}{h_n}) \ge \exp(-1) \sum_{i=1}^{n} \mathbb{1}_{X_i \in A_1},$$

and $\mathbb{1}_{X_i \in A_1}$ follows a Bernoulli distribution with success probability $q_i \ge \frac{1}{L} n h_n^\beta$ according to Assumption 1 (ii). We can apply Chernoff Bounds to lower bound $B(x_0)$:

$$\mathbb{P}\left\{\sum_{i=1}^{n} \mathbb{1}_{X_i \in A_1} \le \frac{nh_n^\beta}{2L}\right\} \le \exp(-\frac{nh_n^\beta}{8L}) \Rightarrow \mathbb{P}\left\{B(x_0) \le \frac{nh_n^\beta}{2eL}\right\} \le \exp(-\frac{nh_n^\beta}{8L}).$$

Using the partitions $\{A_j\}$ and Assumption 1 (i):

$$\begin{aligned}
\Delta(x_0) &\le L \sum_{i=1}^{n} d(x_0, X_i) \exp(-\frac{d(x_0, X_i)}{h_n}), \\
&\le L h_n \exp(1) \sum_{k=1}^{\infty} k \sum_{i: X_i \in A_k} \exp(-k) \\
&\le \min_{k_0} \left\{ L \exp(1) k_0 h_n \sum_{k \le k_0} \sum_{i: X_i \in A_k} H(x_0, X_i) + L h_n \exp(1) \sum_{k > k_0} \sum_{i: X_i \in A_k} k \exp(-k) \right\} \\
&\le \min_{k_0} \left\{ eLk_0 h_n B(x_0) + eLh_n k_0 \exp(-k_0) n \right\} \\
&\le eL\beta \lceil \ln h_n^{-1} \rceil h_n \left(B(x_0) + n h_n^\beta \right),
\end{aligned} \tag{C.3}$$

where we have taken $k_0 = \beta \lceil \ln h_n^{-1} \rceil$ at the last step. Hence, there exists a constant $C > 0$ such that

$$\frac{\Delta(x_0)}{B(x_0) \vee (nh_n^\beta)} \le 2eL\beta \lceil \ln h_n^{-1} \rceil h_n \le C \ln(h_n^{-1}) h_n, \text{ for all } x_0 \in [0,1]^p.$$

Part (ii): Set $Z_{ij} = \frac{H_{ij}}{B_i} \left(\mathbb{1}_{V_j < V_i} - P_{V|X_j}(V_i) \right)$, and $R_i = \sum_{j \ne i} Z_{ij}$. By Hoeffding's lemma, the centered variable Z_{ij} is sub-Gaussian with parameter $\nu_{ij} = \frac{H_{ij}}{2B_i}$ for all i, j and V_i: for all $j \ne i$,

$$\mathbb{E}[\exp(\lambda Z_{ij}) | V_i] \le \exp(\frac{\nu_{ij}^2 \lambda}{2}), \quad \text{for all } \lambda \in \mathbb{R}.$$

Hence, the weighted sum R_i is sub-Gaussian with parameter $\nu_i = \sqrt{\sum_{j \ne i, j \le n} \frac{H_{ij}^2}{4B_i^2}} \le \sqrt{\frac{1}{4B_i}}$, where we have used the fact that $H_{ij} \le 1$ and $B_i = \sum_{j=1}^{n+1} H_{ij}$. Combining it with the sub-Gaussian concentration results, we obtain that

$$\mathbb{P}\left\{|R_i| \ge t | \mathcal{X}, V_i\right\} \le 2\exp(-\frac{t^2}{2\nu_i^2}) \le 2\exp(-2t^2 B_i), \text{ for all } V_i, i = 1, \dots, n+1.$$

Take $t = \sqrt{\frac{\ln n}{B_i}}$, we obtain the desired bound. $\square$

D. Choice of H

D.1. Estimation of the default distance

Let $\mathcal{V}$ be the CV fold partitioning when learning V. We will estimate the spread by learning $|V_i|$ for V_i from the cross-validation step and $i = 1, \ldots, n_0$:

$$V_i \leftarrow \hat{V}^{-i}(X_i^0, Y_i^0),$$

where $\hat{V}^{-i}$ is the score function learned using samples excluding i.

The spread learning step is using the same CV partitioning $\mathcal{V}$. To learn the spread $\rho(X)$, we consider minimizing the MSE with the response $\log(|V_i| + \overline{|V_i|})$, with $\overline{|V_i|}$ be the mean absolute value for V_i across samples in $\mathcal{D}_0 = \{Z_i^0 = (X_i^0, Y_i^0), i = 1, \ldots, n_0\}$. This additional term $\overline{|V_i|}$ is added to reduce the influence of samples with very small empirical $|V_i|$.

We do not claim that learning $\rho(X)$ in such a way is always a good choice. It is a reasonable choice for the regression score. However, for quantile regression score, $|V_i|$ can be large around regions with severe under-coverage or over-coverage, making it a poor target. Despite this, the resulting LCP is similar to CP with a poorly chosen $\hat{\rho}(x)$ for the quantile regression score in our empirical studies.

Our estimated $\hat{\rho}$ is defined as $\hat{\rho} = \exp(\hat{f}(x))$ where $\hat{f}(x)$ is the estimated function from the learning step. We let $\rho_i = \hat{\rho}^{-i}(X_i^0)$ be the estimated spread from the cross-validation step. Let $J \in \mathbb{R}^{n_0 \times p}$ be the Jacobian matrix with $J_{i,} = \frac{\partial \hat{f}^{-i}(X_i^0)}{\partial X_i^0}$. Let $u_{\|} \in \mathbb{R}^{p \times p_0}$ and $u_{\perp} \in \mathbb{R}^{p \times (p - p_0)}$ be the top p_0 and the remaining right singular vectors, with p_0 be a small constant. By default, $p_0 = 1$. We form the projection matrix $P_{\|}$ and $P_{\perp}$ with $u_{\|}$ and $u_{\perp}$:

$$P_{\|} = u_{\|} u_{\|}^\top, \; P_{\perp} = u_{\perp} u_{\perp}^\top.$$

The final dissimilarity measure $d(x_1, x_2)$ is a weighted sum of the three components, and

$$d(x_1, x_2) = \frac{d_1(x_1, x_2)}{\sigma_2} + \frac{(\omega d_2(x_1, x_2) + (1 - \omega) d_3(x_1, x_2))}{\sigma_1},$$

where $d_2(x_1, x_2)$, $d_3(x_1, x_2)$ are projected distances onto $P_{\|}$ and $P_{\perp}$, and $d_1(x_1, x_2)$ are distance in the space of the learned spreading function $\hat{\rho}(x_1), \hat{\rho}(x_2)$ as described in Section 3.3:

- $d_1(x_1, x_2) = \|\hat{\rho}(x_1) - \hat{\rho}(x_2)\|_2$.
- $d_2(x_1, x_2) = \|P_{\|}(x_1 - x_2)\|_2$.
- $d_3(x_1, x_2) = \|P_{\perp}(x_1 - x_2)\|_2$.

We set ω and σ_1, σ_2 as following:

- Let $\mu_{\|}/\mu_{\perp}$ be the mean of $d_2(X_i^0, X_j^0)$ or $d_3(X_i^0, X_j^0)$ for $i \neq j$, then we let $w = \frac{\mu_{\perp}}{\mu_{\perp} + \mu_{\|}}$.
- We let σ_1 be the mean of $\left(\omega d_2(X_i^0, X_j^0) + (1 - \omega) d_3(X_i^0, X_j^0)\right)$ and σ_2 be that mean of $d_1(X_i^0, X_j^0)$, using all pairs $i \neq j$ from $\mathcal{D}_0$.

D.2. Empirical estimate of the objective

We want to minimize a penalized average length of finite PIs:

$$J(h) = \text{Average PI}^{finite} \text{ length} + \lambda \times \text{Average conditional PI}^{finite} \text{ length variability}$$
$$s.t. \; \mathbb{P}(\text{Infinite PI}) \leq \delta.$$

Let $\mathbb{E}_X f(X)$ denote the expectation of some function $f(.)$ with over X. In this tuning section, we consider two specific types of $V(.)$: the scaled regression score and the scaled quantile score, and

$$V(X, Y) = \frac{1}{\sigma(X)} |Y - f(X)|,$$

or

$$V(X, Y) = \frac{1}{\sigma(X)} \max\{q_{lo}(X) - Y, Y - q_{hi}(X)\}.$$

These two score classes will include the four scores considered in our numerical experiments. Let k^* be the selected index from Lemma 2, for this two classes of scores, the PI of Y_{n+1} is constructed as

$$C(X_{n+1}) = [f(X_{n+1}) - \sigma(X_{n+1})\overline{V}_{k^*}, f(X_{n+1}) + \sigma(X_{n+1})\overline{V}_{k^*}],$$

or

$$C(X_{n+1}) = [q_{lo}(X_{n+1}) - \sigma(X_{n+1})\overline{V}_{k^*}, q_{hi}(X_{n+1}) + \sigma(X_{n+1})\overline{V}_{k^*}].$$

In both cases, the length over the constructed PI of Y_{n+1} is additive on $\sigma(X_{n+1})\overline{V}_{k^*}$, and hence, minimizing PI of Y_{n+1} is equivalent to minimizing $\sigma(X_{n+1})\overline{V}_{k^*}$, and the conditional variability of the PI is the same as the variability of $\sigma(X_{n+1})\overline{V}_{k^*}$ conditional on X_{n+1}. Hence, after omitting components that do not depend on h, we can express the terms in the above objective as

- Average PI^{finite} length: $\mathbb{E}_{Z_{1:n},X_{n+1}}\left[\sigma(X_{n+1})\overline{V}_{k^*}|k^* \leq n\right]$. It depends on $Z_{1:n}, X_{n+1}$ as well as the tuning parameter h. (Recall that when $k^* = n+1$, $\overline{V}_{n+1} = \infty$.)
- Average conditional PI^{finite} length variability: $\sqrt{\mathbb{E}_{Z_{1:n},X_{n+1}}\left[\sigma(X_{n+1})^2\left(\overline{V}_{k^*} - \mu(X_{n+1})\right)^2 |k^* \leq n\right]}$, where $\mu(X_{n+1}) = \mathbb{E}_{Z_{1:n}}\left[\overline{V}_{k^*}|k^* \leq n, X_{n+1}\right]$ is the average length finite PI at X_{n+1}, marginalized over $Z_{1:n}$.
- Average percent of infinite PI: $\mathbb{P}(k^* = n+1)$.

We estimate the above quantities with empirical estimates using $\mathcal{D}_0$. As in the previous section, we consider the case where the function form $V(X,Y)$ is estimated by CV and $V_i \leftarrow \hat{V}^{-i}(X_i^0, Y_i^0)$. For example, we want to construct the score function $V(X,Y) = |Y - f(X)|$ where $f(X)$ is the mean prediction function. Then, $\hat{V}^{-i}(X_i^0, Y_i^0)$ is calculated as

$$\hat{V}^{-i}(X_i^0, Y_i^0) = |Y_i^0 - \hat{f}^{-k}(x)|,$$

where $\hat{f}^{-k}(.)$ is the learned mean function using data excluding fold k that includes sample i. We also estimate the spreads and define the distance on $\mathcal{D}_0$ using the CV estimates.

Given the dissimilarity measure d_{ij} for any pair (X_i^0, X_j^0), and thus $H_{ij} = \exp(-\frac{d_{ij}}{h})$ for a given h, we estimate the empirical loss for $h \in \{h_1, \ldots, h_m\}$ as below:

- Estimation of average length and infinite PI probability:
  - We subsample $\tilde{n} = (n+1) \wedge n_0$ samples without replacement from $\mathcal{D}_0$, let the set be $\mathcal{S}$ and construct PI for each sample $i \in \mathcal{S}$ with a calibration set $\mathcal{S} \setminus \{i\}$. Let L_i be the scaled length for the constructed PI (scaled by $\sigma(X_i^0)$).
  - The probability of having infinite PI is estimated as $C_1(h) = \frac{\#\{i\in\mathcal{S}, L_i=\infty\}}{\tilde{n}}$, and the average finite PI length is estimated as $C_2(h) = \frac{\sum_{i\in\mathcal{S}, L_i<\infty} L_i}{\#\{b:L_{ib}<\infty\}\vee 1}$.

  The above estimates can be repeated for multiple times when n_0 is much larger than $(n+1)$.
- Estimation of conditional variability:
  - Repeat B times the PI construction: for $b = 1, \ldots, B$, we subsample n samples with replacement from $\mathcal{D}_0$, and let the length of scaled PI of V at Z_i^0 be L_{ib} for $i = 1, \ldots, n_0$.
  - Calculate the finite conditional mean as $\mu_i = \frac{\sum_{b:L_{ib}<\infty} L_{ib}}{\#\{b:L_{ib}<\infty\}\vee 1}$.
  - Calculate the conditional variance as $s_i = \frac{\sum_{b:L_{ib}<\infty}(L_{ib}-\mu_i)^2}{\#\{b:L_{ib}<\infty\}\vee 1}$.
  - The average conditional variability for PI with finite length is estimated as $C_{3,h} = \sqrt{\frac{\sum_i(\#\{b:L_{ib}<\infty\}\times s_i)}{\#\{(i,b):L_{ib}<\infty\}}}$

We take h from the candidate set to minimize the empirical objective:

$$h = \arg\min_{C_1(h)\leq\delta}\left(C_2(h) + \lambda C_3(h)\right).$$

REFERENCES

BARBER, R. F., CANDES, E. J., RAMDAS, A. & TIBSHIRANI, R. J. (2019). Conformal prediction under covariate shift. *arXiv:1904.06019* .