A. TECHNICAL PROOFS

A.1 Proofs of Section 2

Proof of Theorem 2.1. The first part (the lower bound) comes directly from the definition of the conformal interval in (7), and the (discrete) p-value property in (6). We focus on the second part (upper bound). Define $\alpha' = \alpha - 1/(n+1)$. By assuming a continuous joint distribution of the fitted residuals, we know that the values $R_{y,1}, \ldots, R_{y,n+1}$ are all distinct with probability one. The set C_{conf} in (7) is equivalent to the set of all points y such that $R_{y,n+1}$ ranks among the $\lceil (n+1)(1-\alpha) \rceil$ smallest of all $R_{y,1}, \ldots, R_{y,n+1}$. Consider now the set $D(X_{n+1})$ consisting of points y such that $R_{y,n+1}$ is among the $\lceil (n+1)\alpha' \rceil$ largest. Then by construction

$$\mathbb{P}\Big(Y_{n+1} \in D(X_{n+1})\Big) \geq \alpha',$$

and yet $C_{\text{conf}}(X_{n+1}) \cap D(X_{n+1}) = \emptyset$, which implies the result. ■

Proof of Theorem 2.2. The first part (lower bound) follows directly by symmetry between the residual at (X_{n+1}, Y_{n+1}) and those at (X_i, Y_i), $i \in \mathcal{I}_2$. We prove the upper bound in the second part. Assuming a continuous joint distribution of residual, and hence no ties, the set $C_{\text{split}}(X_{n+1})$ excludes the values of y such that $|y - \widehat{\mu}(X_{n+1})|$ is among the $(n/2) - \lceil (n/2+1)(1-\alpha) \rceil$ largest in $\{R_i : i \in \mathcal{I}_2\}$. Denote the set of these excluded points as $D(X_{n+1})$. Then again by symmetry,

$$\mathbb{P}\Big(Y_{n+1} \in D(X_{n+1})\Big) \geq \frac{(n/2) - \lceil (n/2+1)(1-\alpha) \rceil}{n/2+1} \geq \alpha - 2/(n+2),$$

which completes the proof. ■

Proof of Theorem 2.4. Without loss of generality, we assume that the sample size is $2n$. The individual split conformal interval has length infinity if $\alpha/N < 1/n$. Therefore, we only need to consider $2 \leq N \leq \alpha n \leq n$. Also in this proof we will ignore all the rounding issues by directly working with the empirical quantiles. The differences caused by rounding is negligible.

For each $1 \leq j \leq N$, $C_{\text{split},j}(X)$, the jth split conformal prediction band at X, is an interval with radius $\widehat{F}_{n,j}^{-1}(1 - \alpha/N)$, where $\widehat{F}_{n,j}$ is the empirical CDF of fitted absolute residuals in the ranking subsample in the jth split.

We focus on the event $\{\sup_{1\le j\le N}\|\widehat{\mu}_j-\mu_0\|_\infty<\eta_n\}$, which has probability at least $1-N\rho_n\ge 1-n\rho_n\to 1$. In this event the length of $C_{\rm split}^{(N)}(X)$ is at least

$$2\min_{1\le j\le N}\widetilde{F}_{n,j}^{-1}(1-\alpha/N)-2\eta_n\,,$$

where $\widetilde{F}_{n,j}$ is the empirical CDF of the absolute residuals in the ranking subsample in the jth split.

Without loss of generality, let $C_{\rm split}(X)=C_{{\rm split},1}(X)$, with length no more than $2\widetilde{F}_{n,1}^{-1}(1-\alpha)+2\eta_n$ on the events we focus on. Therefore, it suffices to show that

$$P\left[\widetilde{F}_{n,1}^{-1}(1-\alpha)<\widetilde{F}_{n,j}^{-1}(1-\alpha/N)-2\eta_n,\ \ \forall\ 1\le j\le N\right]\to 1\,. \tag{16}$$

Let $\widetilde{F}$ be the CDF of $|Y-\mu_0(X)|$. Note that it is $\widetilde{F}_j$, instead of $\widehat{F}_j$, that corresponds to F. By Dvoretzky-Kiefer-Wolfowitz inequality we have,

$$\begin{aligned}P\left[\widetilde{F}_{n,j}^{-1}(1-\alpha/N)\le\widetilde{F}^{-1}(1-\alpha/1.6)\right]&\le P\left[\|\widetilde{F}_{n,j}-\widetilde{F}\|_\infty\ge\alpha(1/1.6-1/N)\right]\\&\le P\left[\|\widetilde{F}_{n,j}-\widetilde{F}\|_\infty\ge\alpha/8\right]\\&\le 2\exp\left(-n\alpha^2/32\right)\end{aligned}$$

Taking union bound

$$P\left[\inf_{1\le j\le N}\widetilde{F}_{n,j}^{-1}(1-\alpha/N)\le\widetilde{F}^{-1}(1-\alpha/1.6)\right]\le 2N\exp(-n\alpha^2/32)\,.$$

On the other hand

$$P\left[\widetilde{F}_{n,1}^{-1}(1-\alpha)\ge\widetilde{F}^{-1}(1-\alpha/1.4)\right]\le P\left[\|\widetilde{F}_{n,1}-\widetilde{F}\|_\infty\ge\alpha(1-1/1.4)\right]\le 2\exp(-n\alpha^2/8)\,.$$

So with probability at least $1-2\exp(-n\alpha^2/8)-2N\exp(-n\alpha^2/32)$ we have

$$\inf_{1\le j\le N}\widetilde{F}_{n,j}^{-1}(1-\alpha/N)-\widetilde{F}_{n,1}^{-1}(1-\alpha)\ge\widetilde{F}^{-1}(1-\alpha/1.6)-\widetilde{F}^{-1}(1-\alpha/1.4)>0\,.$$

Therefore we conclude (16) because $\eta_n=o(1)$. ■

Proof of Theorem 2.3. Comparing the close similarity of $\widetilde{d}_1$ in (12) and d in Algorithm 2, we see that $\widetilde{d}_1=d$ if we choose the target coverage levels to be $1-\alpha$ for the regular split conformal band $C_{\rm split}$, and $1-(\alpha+2\alpha/n)$ for the modified ROO split conformal band $\widetilde{C}_{\rm roo}$. The desired result follows immediately by replacing α by $\alpha+2\alpha/n$ in Theorem 5.1, as it applies to $\widetilde{C}_{\rm roo}$ (explained in the above remark). ■

A.2 Proofs of Section 5

Proof of Theorem 5.1. For notation simplicity we assume $I_1 = \{1, ..., n/2\}$, and r_i's are in increasing order for $i = 1, ..., n/2$. Let $m = \lceil (1-\alpha)n/2 \rceil$. Then $\mathbb{1}(Y_i \in C_{\rm roo}(X_i)) = \mathbf{1}(r_i \le d_i)$ where d_i is the mth smallest value in $r_1, ..., r_{i-1}, r_{i+1}, ..., r_{n/2}$. Now we consider changing a sample point, say, (X_j, Y_j), in $\mathcal{I}_1$ and denote the resulting possibly disordered residuals by $r_1', ..., r_{n/2}'$, and define d_i' correspondingly. We now consider the question "for which values of $i \in \mathcal{I}_1 \backslash \{j\}$ can we have $\mathbb{1}(r_i \le d_i) \neq \mathbb{1}(r_i' \le d_i')$?" Recall that we assume $r_1 \le r_2 \le ... \le r_{n/2}$. If $i \le m-1$ and $i \neq j$, then $d_i \ge r_m$, $d_i' \ge r_{m-1}$, $r_i = r_i'$, and hence $\mathbb{1}(r_i \le d_i) = \mathbb{1}(r_i' \le d_i') = 1$. If $i \ge m+2$ and $i \neq j$, then using similar reasoning we have $\mathbb{1}(r_i \le d_i) = \mathbb{1}(r_i' \le d_i') = 0$. Therefore, changing a single data point can change $\mathbb{1}(Y_i \in C_{\rm roo}(X_i))$ for at most three values of i ($i = m, m+1, j$). Because the input sample points are independent, using McDiarmid's inequality we have

$$\mathbb{P}\left((2/n)\sum_{i\in\mathcal{I}_1} \mathbb{1}(Y_i \in C_{\rm roo}(X_i)) \le 1-\alpha-\epsilon\right) \le \exp(-cn\epsilon^2)\,.$$

The claim follows by switching $\mathcal{I}_1$ and $\mathcal{I}_2$ and adding the two inequalities up.

Now we consider the other direction. We only need to show that $\mathbb{P}(Y_j \notin C_{\rm roo}(X_j)) \ge \alpha - 2/n$. Under the continuity assumption, with probability one the residuals r_j are all distinct. Let $j \in \mathcal{I}_k$ for $k = 1$ or 2. By construction, $C_{\rm roo}(X_j)$ does not contain the y values such that $|y - \widehat{\mu}_{3-k}(X_j)|$ is among the $n/2 - \lceil (n/2)(1-\alpha) \rceil$ largest of $\{r_i : i \in \mathcal{I}_k \backslash \{j\}\}$. Denote this set by $D_{\rm roo}(X_j)$. Then the standard conformal argument implies that

$$\mathbb{P}(Y_i \in D_{\rm roo}(X_i)) \ge \frac{n/2 - \lceil (n/2)(1-\alpha) \rceil}{n/2} \ge \alpha - \frac{2}{n}\,.$$

And we can establish the corresponding exponential deviation inequality using the same reasoning above.

For $\widetilde{C}_{\rm roo}(X_j)$, the lower bound follows from that of $C_{\rm roo}(X_j)$ because $\widetilde{C}_{\rm roo}(X_j) \supseteq C_{\rm roo}(X_j)$. To prove the upper bound, realize that $\widetilde{C}_{\rm roo}(X_j)$ does not contain y such that $|y - \widehat{\mu}_{3-k}(X_j)|$ is among the $(n/2) - \lceil (n/2)(1-\alpha) \rceil - 1$ largest of $\{r_i : i \in \mathcal{I}\}$. Hence it does not contain all y's such that $|y - \widehat{\mu}_{3-k}(X_j)|$ is among the $(n/2) - \lceil (n/2)(1-\alpha) \rceil - 2$ largest of $\{r_i : i \in \mathcal{I} \backslash \{j\}\}$. Comparing with the argument for $C_{\rm roo}$, the extra -2 in the ranking changes $2/n$ to $6/n$ in the second probability statement in the theorem. ∎

A.3 Proofs of Section 3

Proof of Theorem 3.1. For any $t > 0$, by Fubini theorem and independence between ϵ and (Δ_n, X),

$$\begin{aligned} F_n(t) &= \mathbb{P}(|Y - \widehat{\mu}_n(X)| \le t) \\ &= \mathbb{P}(-t + \Delta_n(X) \le \epsilon \le t + \Delta_n(X)) \\ &= \mathbb{E}_{\widehat{\mu}_n, X}[F_0(t + \Delta_n(X)) - F_0(-t + \Delta_n(X))], \end{aligned} \tag{17}$$

where F_0 is the CDF of ϵ.

Let f_0 be the density function of F_0. We can approximate F_0 at any t using first order Taylor expansion

$$F_0(t + \delta) = F_0(t) + \delta f_0(t) + \delta^2 R(t, \delta)\,,$$

where $R(t, \delta) = 0.5 \int_0^1 (1 - u) f_0'(t + u\delta) du$ satisfies $\sup_{t,\delta} |R(t, \delta)| \le M/4$.

Next, using symmetry of F_0 we have $f_0(t) = f_0(-t)$ for all t, the RHS of (17) becomes

$$\begin{aligned} &\mathbb{E}_{\widehat{\mu}_n, X}[F_0(t + \Delta_n(X)) - F_0(-t + \Delta_n(X))] \\ =&\mathbb{E}_{\widehat{\mu}_n, X}\left[F_0(t) + \Delta_n(X) f_0(t) + \Delta_n^2(X) R(t, \Delta_n(X)) \right. \\ &\qquad \left. - F_0(-t) - \Delta_n(X) f_0(-t) - \Delta_n^2(X) R(-t, \Delta_n(X))\right] \\ =&F_0(t) - F_0(-t) + \mathbb{E}_{\widehat{\mu}_n, X}[\Delta_n^2(X) W] \\ =&F(t) + \mathbb{E}_{\widehat{\mu}_n, X}[\Delta_n^2(X) W] \end{aligned}$$

where $W = R(t, \Delta_n(X)) - R(-t, \Delta_n(X))$. Equation (10) follows immediately since $|W| \le M/2$, almost surely.

Next we show equation (11). Because F has density at least $r > 0$ in an open neighborhood of q_α. If $t < q_\alpha - \delta$ for some $\delta > (M/2r)\mathbb{E}(\Delta_n^2(X))$ then

$$\begin{aligned} F_q(t) \le& F(q_\alpha - \delta) + (M/2)\mathbb{E}(\Delta_n^2(X)) \\ \le& F(q_\alpha) - \delta r + (M/2)\mathbb{E}(\Delta_n^2(X)) \\ <&1 - \alpha\,. \end{aligned}$$

Thus $q_{n,\alpha} \ge q_\alpha - (M/2r)\mathbb{E}(\Delta_n^2(X))$. Similarly we can show that $q_{n,\alpha} \le q_\alpha + (M/2r)\mathbb{E}(\Delta_n^2(X))$, and hence establish the claimed result. ∎

Proof of Theorem 3.2. Let $\widetilde{q}_\alpha$ be the upper α quantile of $|Y-\mu_0(X)|$. We first show that

$$|\widetilde{q}_\alpha - q_{n,\alpha}| \le \rho_n/r + \eta_n \,. \tag{18}$$

To see this, note that

$$\begin{aligned}
&P(|Y-\widehat{\mu}_n(X)| \le \widetilde{q}_{\alpha+\rho_n} - \eta_n) \\
\le& P\left[|Y-\widehat{\mu}_n(X)| \le \widetilde{q}_{\alpha+\rho_n} - \eta_n, \|\mu_0 - \widehat{\mu}_n\|_\infty \le \eta_n\right] + \rho_n \\
\le& \mathbb{P}\left[|Y-\mu_0(X)| \le \widetilde{q}_{\alpha+\rho_n}\right] + \rho_n \\
=& 1-\alpha-\rho_n+\rho_n = 1-\alpha\,.
\end{aligned}$$

Thus $q_{n,\alpha} \ge \widetilde{q}_{\alpha+\rho_n} - \eta_n \ge \widetilde{q}_\alpha - \rho_n/r - \eta_n$. Similarly, we have $q_{n,\alpha} \le \widetilde{q}_{\alpha-\rho_n} + \eta_n \le \widetilde{q}_\alpha + \rho_n/r + \eta_n$.

The width of split conformal band equals $2\widehat{F}_n^{-1}(1-\alpha)$, where $\widehat{F}_n$ is the empirical CDF of $\{|Y_i - \widehat{\mu}_n(X_i)| : 1 \le i \le n\}$, and $\widehat{\mu}_n = \mathcal{A}_n((X_i,Y_i)_{i=n+1}^{2n})$. On the event $\{\|\widehat{\mu}_n - \mu_0\|_\infty \le \eta_n\}$, we have $\big||Y_i - \widehat{\mu}_n(X_i)| - |Y_i - \mu_0(X_i)|\big| \le \eta_n$ for all $1 \le i \le n$. Therefore, let $\widetilde{F}_n$ be the empirical CDF of $\{|Y_i - \mu_0(X_i)| : 1 \le i \le n\}$, we have

$$\mathbb{P}\left(|\widehat{F}_n^{-1}(1-\alpha) - \widetilde{F}_n^{-1}(1-\alpha)| \le \eta_n\right) \ge 1-\rho_n\,. \tag{19}$$

Using standard empirical quantile theory for i.i.d data and using the assumption that $\widetilde{f}$ is bounded from below by $r>0$ in a neighborhood of its upper α quantile, we have

$$\widetilde{F}_n^{-1}(1-\alpha) = \widetilde{q}_\alpha + O_P(n^{-1/2}). \tag{20}$$

Combining (18), (19), and (20), we conclude that

$$|\widehat{F}_n^{-1}(1-\alpha) - q_{n,\alpha}| = O_P(\eta_n + \rho_n + n^{-1/2})\,. \qquad \blacksquare$$

Proof of Theorem 3.3. We focus on the event

$$\{\|\widehat{\mu}_n - \mu_0\|_\infty \le \eta_n\} \cap \left\{\sup_{y\in\mathcal{Y}} \|\widehat{\mu}_n - \widehat{\mu}_{n,(X,y)}\|_\infty \le \eta_n\right\},$$

which, by assumption, has probability at least $1-2\rho_n \to 1$.

On this event, we have

$$\left||Y_i - \widehat{\mu}_{n,(X,y)}(X_i)| - |Y_i - \mu_0(X_i)|\right| \le 2\eta_n\,, \quad i=1,...,n\,. \tag{21}$$

$$\Big||y-\widehat{\mu}_{n,(X,y)}(X)|-|y-\mu_0(X)|\Big|\le 2\eta_n\,. \tag{22}$$

With (21) and (22), by the construction of full conformal prediction interval we can directly verify the following two facts.

1. $y\in C_{\rm conf}(X)$ if $|y-\mu_0(X)|\le \widetilde{F}_n^{-1}(1-\alpha)-4\eta_n$;
2. $y\notin C_{\rm conf}(X)$ if $|y-\mu_0(X)|\ge \widetilde{F}_n^{-1}(1-(\alpha-3/n))+4\eta_n$,

where $\widetilde{F}_n$ is the empirical CDF of $\{|Y_i-\mu_0(X_i)|:1\le i\le n\}$.

Therefore, the length of $C_{\rm conf}(X)$ satisfies

$$\nu_{n,{\rm conf}}(X)=2\widetilde{q}_\alpha+O_P(\eta_n+n^{-1/2})\,.$$

The claimed result follows by further combining the above equation with (18). ■

Proof of Theorem 3.4. Without loss of generality, we assume that $C_{\rm split}(\cdot)$ is obtained using $2n$ data points. The proof consists of two steps. First we show that $\widehat{\mu}_n(X)-\mu(X)=o_P(1)$. Second we show that $\widehat{F}_n^{-1}(1-\alpha)-q_\alpha=o_P(1)$, where $\widehat{F}_n$ is the empirical CDF of $\{|Y_i-\widehat{\mu}_n(X_i)|:1\le i\le n\}$ with $\widehat{\mu}_n=\mathcal{A}_n((X_i,Y_i)_{i=n+1}^{2n})$.

We now show the first part. Throughout this proof we focus on the event that $\mathbb{E}_X(\widehat{\mu}_n(X)-\mu(X))^2\le\eta_n$, which has probability at least $1-\rho_n$ by Assumption A4. On this event, using Markov's inequality, we have that $\mathbb{P}(X\in B_n^c\mid\widehat{\mu}_n)\ge 1-\eta_n^{1/3}$, where $B_n=\{x:|\widehat{\mu}_n(x)-\mu(x)|\ge\eta_n^{1/3}\}$. Therefore we conclude that $\mathbb{P}_{X,\widehat{\mu}_n}(|\widehat{\mu}_n(X)-\mu(X)|\ge\eta_n^{1/3})\le\eta_n^{1/3}+\rho_n\to 0$ as $n\to\infty$. Part 1 of the proof is complete.

For the second part, define $\mathcal{I}_1=\{i:\ 1\le i\le n,\ X_i\in B_n^c\}$ and $\mathcal{I}_2=\{1,...,n\}\backslash\mathcal{I}_1$. Note that B_n is independent of $(X_i,Y_i)_{i=1}^n$. Using Hoeffding's inequality conditionally on $\widehat{\mu}_n$ we have $|\mathcal{I}_2|\le n\eta_n^{1/3}+c\sqrt{n\log n}=o(n)$ with probability tending to 1, for some absolute constant $c>0$. The probability holds both conditionally or unconditionally on $\widehat{\mu}_n$.

Let $\widehat{G}_{n,1}$ be the empirical CDF of $\{|Y_i-\widehat{\mu}_n(X_i)|:i\in\mathcal{I}_1\}$, and $\widetilde{G}_{n,1}$ be the empirical CDF of $\{|Y_i-\mu(X_i)|:i\in\mathcal{I}_1\}$. By definition of $\mathcal{I}_1$ we know that $\big||Y_i-\widehat{\mu}_n(X_i)|-|Y_i-\mu(X_i)|\big|\le\eta_n^{1/3}$

for all $i \in \mathcal{I}_1$. All empirical quantiles of $\widehat{G}_{n,1}$ and $\widetilde{G}_{n,1}$ are at most $O_P(\sqrt{n})$ apart, because $|\mathcal{I}_1| = n(1 + o_P(1))$.

The half width of $C_{\text{split}}(X)$ is $\widehat{F}_n^{-1}(1-\alpha)$. According to the definition of $\mathcal{I}_1$, we have

$$\widehat{G}_{n,1}^{-1}\left(1 - \frac{n\alpha}{|\mathcal{I}_1|}\right) \le \widehat{F}_n^{-1}(1-\alpha) \le \widehat{G}_{n,1}\left(1 - \frac{n\alpha - |\mathcal{I}_2|}{|\mathcal{I}_1|}\right).$$

Both $\frac{n\alpha}{|\mathcal{I}_1|}$ and $\frac{n\alpha - |\mathcal{I}_2|}{|\mathcal{I}_1|}$ are $\alpha + o_P(1)$. As a result we conclude that

$$\widehat{F}_n^{-1}(1-\alpha) - q_\alpha = o_P(1).$$

The second part of the proof is complete. ■

Proof of Theorem 3.5. The proof naturally combines those of Theorems 3.3 and 3.4.

Using the same argument as in the proof of Theorem 3.4, we can define the set B_n and index sets $\mathcal{I}_1$, $\mathcal{I}_2$.

Now we assume the event $\{X \in B_n^c\}$, which has probability tending to 1. Then by definition of B_n and the fact that $\eta_n \le \eta_n^{1/3}$ we have

$$\Big| |Y_i - \widehat{\mu}_{n,(X,y)}(X_i)| - |Y_i - \mu_0(X_i)| \Big| \le 2\eta_n^{1/3}, \quad \forall\, i \in \mathcal{I}_1. \tag{23}$$

$$\Big| |y - \widehat{\mu}_{n,(X,y)}(X)| - |y - \mu_0(X)| \Big| \le 2\eta_n^{1/3}. \tag{24}$$

By definition of $C_{\text{conf}}(X)$ and following the same reasoning as in the proof of Theorem 3.3, we can verify the following facts:

1. $y \in C_{\text{conf}}(X)$ if $|y - \mu_0(X)| \le \widetilde{G}_{n,1}^{-1}\left(1 - \frac{n\alpha}{|\mathcal{I}_1|}\right) - 4\eta_n^{1/3}$;
2. $y \notin C_{\text{conf}}(X)$ if $|y - \mu_0(X)| \ge \widetilde{G}_{n,1}^{-1}\left(1 - \frac{n\alpha - |\mathcal{I}_2| - 3}{|\mathcal{I}_1|}\right) + 4\eta_n^{1/3}$,

where $\widetilde{G}_{n,1}$ is the empirical CDF of $\{|Y_i - \mu_0(X_i)| : i \in \mathcal{I}_1\}$.

Both $\frac{n\alpha}{|\mathcal{I}_1|}$ and $\frac{n\alpha - |\mathcal{I}_2| - 3}{|\mathcal{I}_1|}$ are $\alpha + o_P(1)$, and hence

$$\widetilde{G}_{n,1}^{-1}\left(1 - \frac{n\alpha}{|\mathcal{I}_1|}\right) = q_\alpha + o_P(1), \quad \widetilde{G}_{n,1}^{-1}\left(1 - \frac{n\alpha - |\mathcal{I}_2| - 3}{|\mathcal{I}_1|}\right) = q_\alpha + o_P(1).$$

Thus the lower (upper) end point of $C_{\text{conf}}(X)$ is $q_\alpha + o_P(1)$ below (above) $\mu(X)$. The proof is complete. ■