Instant-by-instant and simultaneous confidence prediction bands

Davidson Lova Razafindrakoto[1, 2]

[1]Laboratoire SAMM, Université Paris 1 Panthéon-Sorbonne, Paris, France.
[2]Safran Aircraft Engines, Paris, France.

davidson-lova.razafindrakoto@safrangroup.com

Appendix

Definition 1. *Let K the size of the fleet and T the size of the time grid be non-zero integers. Let a cumulative event number evolution curve $\xi = (\xi_1, \ldots, \xi_T)$ be a random vector where for every $\tau \in \{1, \ldots, T\}$, ξ_τ is a $\{1, \ldots, K\}$-valued random variable. Let a band $B = (B_1, \ldots, B_T)$ be such that for every $\tau \in \{1, \ldots, K\}, \hat{B}_{\alpha;\tau} \subseteq \{1, \ldots, K\}$.*

Definition 2 (Instant-by-instant coverage)**.** *Let $\alpha \in (0, 1)$. A band B controls the instant-by-instant coverage of the curve ξ at a control level α if*

$$\forall \tau \in \{1, \ldots, T\}, \quad \mathbb{P}\left[\xi_\tau \in \hat{B}_{\alpha;\tau}\right] \geq 1 - \alpha.$$

Definition 3 (γ-Simultaneous coverage)**.** *Let $\alpha \in (0, 1)$ and $\gamma \in [0, 1]$. A band B controls the γ-simultaneous coverage of the curve ξ at a control level α if*

$$\mathbb{P}\left[\frac{\operatorname{Card}\left(\left\{\tau \in \{1, \ldots, T\} : \xi_\tau \in \hat{B}_{\alpha;\tau}\right\}\right)}{T} \geq 1 - \gamma\right] \geq 1 - \alpha.$$

Proposition 1 (Instant-by-instant prediction band, MD-full)**.** *Let $\xi_1, \ldots, \xi_n$ be n independent copies of ξ. For any control level $\alpha \in [0, 1]$, the band $\hat{B}_\alpha^{\mathrm{MD-full}}$ defined as, for every $\tau \in \{1, \ldots, T\}$*

$$\hat{B}_{\alpha;\tau}^{\mathrm{MD-full}} := \left\{k \in \{1, \ldots, K\} : \hat{p}_\tau(k) \geq \hat{\ell}_{\alpha;\tau} - \frac{1}{n}\right\} \tag{1}$$

ensures control of the instant-by-instant coverage of ξ at a control level α, where the marginal empirical probability density function $\hat{p}_\tau$ is defined in Equation (Eq. (2)) and the threshold $\hat{\ell}_{\alpha;\tau}$ in Equation (Eq. (3)).

Proof of Proposition 1. Let $\alpha \in [0, 1]$ and $\tau \in \{1, \ldots, T\}$. Following Lei, Robins, and Wasserman (2013), a full conformal prediction $\hat{C}_{\alpha;\tau}$ region is built form the observations $\xi_{1;\tau}, \ldots, \xi_{n;\tau}$ for ξ_τ as follows

$$\hat{C}_{\alpha;\tau} = \{k \in \{1, \ldots, K\} : \hat{\pi}_\tau(k) > \alpha\}$$

where for every $k \in \{1, \ldots, K\}$

$$\hat{\pi}_\tau(k) := \frac{1 + \sum_{i=1}^n \mathbb{1}\left\{\hat{p}_\tau^k(\xi_{i;\tau}) \leq \hat{p}_\tau^k(k)\right\}}{n+1}$$

where for every $l \in \{1, \ldots, K\}$

$$\hat{p}_\tau^k(l) = \frac{n}{n+1}\hat{p}_\tau(l) + \frac{1}{n+1}\mathbb{1}\left\{k = l\right\},$$

and the empirical marginal probability density function is given by

$$\hat{p}_\tau(l) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{\xi_{i;\tau} = l\right\} \tag{2}$$

Since $\xi_{1;\tau}, \ldots, \xi_{n;\tau}$ are i.i.d. copies of ξ_τ, $\hat{C}_{\alpha;\tau}$ is a confidence prediction region (Vovk, Gammerman, and Shafer, 2005)

$$\mathbb{P}\left[\xi_\tau \in \hat{C}_{\alpha;\tau}\right] \geq 1 - \alpha.$$

Let us now provide an explicit expression of this region.

$$\begin{aligned} k \in \hat{C}_{\alpha;\tau} &\iff \hat{\pi}_\tau\left(k\right) > \alpha \\ &\iff \frac{1 + \sum_{i=1}^{n} \mathbb{1}\left\{\hat{p}_\tau^k\left(\xi_{i;\tau}\right) \leq \hat{p}_\tau^k(k)\right\}}{n+1} > \alpha \\ &\iff \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{\hat{p}_\tau^k\left(\xi_{i;\tau}\right) > \hat{p}_\tau^k(k)\right\} < (1-\alpha)\left(1 + \frac{1}{n}\right). \end{aligned}$$

Let us consider the term on the l.h.s.

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{\hat{p}_\tau^k\left(\xi_{i;\tau}\right) > \hat{p}_\tau^k(k)\right\} \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{K} \mathbb{1}\left\{l = \xi_{i;\tau}\right\} \mathbb{1}\left\{\hat{p}_\tau^k\left(\xi_{i;\tau}\right) > \hat{p}_\tau^k(k)\right\} \\ &= \sum_{l=1}^{K} \hat{p}_\tau(l)\mathbb{1}\left\{\hat{p}_\tau^k(l) > \hat{p}_\tau^k(k)\right\} \\ &= \sum_{l=1}^{K} \hat{p}_\tau(l)\mathbb{1}\left\{\frac{n}{n+1}\hat{p}_\tau(l) + \frac{1}{n+1}\mathbb{1}\left\{k = l\right\} > \frac{n}{n+1}\hat{p}_\tau(k) + \frac{1}{n+1}\right\} \\ &= \sum_{l=1}^{K} \hat{p}_\tau(l)\mathbb{1}\left\{\hat{p}_\tau(l) + \frac{1}{n}\mathbb{1}\left\{k = l\right\} > \hat{p}_\tau(k) + \frac{1}{n}\right\} \\ &= \hat{p}_\tau(k)\mathbb{1}\left\{\hat{p}_\tau(k) + \frac{1}{n} > \hat{p}_\tau(k) + \frac{1}{n}\right\} + \sum_{l=1, l\neq k}^{K} \hat{p}_\tau(l)\mathbb{1}\left\{\hat{p}_\tau(l) > \hat{p}_\tau(k) + \frac{1}{n}\right\} \\ &= \hat{p}_\tau(k)\mathbb{1}\left\{\hat{p}_\tau(k) > \hat{p}_\tau(k) + \frac{1}{n}\right\} + \sum_{l=1, l\neq k}^{K} \hat{p}_\tau(l)\mathbb{1}\left\{\hat{p}_\tau(l) > \hat{p}_\tau(k) + \frac{1}{n}\right\} \\ &= \sum_{l=1}^{K} \hat{p}_\tau(l)\mathbb{1}\left\{\hat{p}_\tau(l) > \hat{p}_\tau(k) + \frac{1}{n}\right\}. \end{aligned}$$

It follows that

$$k \in \hat{C}_{\alpha;\tau} \iff \sum_{l=1}^{K} \hat{p}_\tau(l)\mathbb{1}\left\{\hat{p}_\tau(l) > \hat{p}_\tau(k) + \frac{1}{n}\right\} < (1-\alpha)\left(1 + \frac{1}{n}\right)$$

$$\iff \hat{p}_\tau(k) \geq \hat{\ell}_{\alpha;\tau},$$

where the threshold $\hat{\ell}_{\alpha;\tau}$ is defined as

$$\hat{\ell}_{\alpha;\tau} := \inf\left\{t \in [0,1] : \sum_{l=1}^{K} \hat{p}_\tau(l)\mathbb{1}\left\{\hat{p}_\tau(l) > t\right\} < (1-\alpha)\left(1+\frac{1}{n}\right)\right\}. \tag{3}$$

How to compute $\hat{\ell}_{\alpha;\tau}$. The function $G : [0,1] \to [0,1]$, $t \mapsto \hat{G}_\tau(t) := \sum_{l=1}^{K} \hat{p}_\tau(l)\mathbb{1}\left\{\hat{p}_\tau(l) > t\right\}$ is decreasing piece-wise constant right-continuous function which changes value at $\hat{p}_\tau(1), \ldots, \hat{p}_\tau(K)$. Let $x_1 > \ldots > x_\kappa$ be the distinct values among $\hat{p}_\tau(1), \ldots, \hat{p}_\tau(K)$ sorted from largest to smallest, and for all $j \in \{1, \ldots, \kappa\}$, $\rho_j := \sum_{l=1}^{K} \hat{p}_\tau(k)\mathbb{1}\left\{\hat{p}_\tau(k) = x_j\right\}$. Consider the cumulative sum of $\rho_1, \ldots, \rho_\kappa$, $s_1, \ldots s_\kappa$. It follows that $G(x_1) = 0$ and for every $j \in \{2, \ldots, \kappa\}$

$$\begin{aligned}
G(x_j) &= \sum_{l=1}^{K} \hat{p}_\tau(l)\mathbb{1}\left\{\hat{p}_\tau(l) > x_j\right\} = \sum_{l=1}^{K}\sum_{i=1}^{\kappa} \mathbb{1}\left\{x_i = \hat{p}_\tau(l)\right\}\hat{p}_\tau(l)\mathbb{1}\left\{\hat{p}_\tau(l) > x_j\right\} \\
&= \sum_{i=1}^{\kappa} \rho_i \mathbb{1}\left\{x_i > x_j\right\} = \sum_{i=1}^{j-1} \rho_i = s_{j-1}.
\end{aligned}$$

It follows that, one only has to check for finite number of values

$$\begin{aligned}
\hat{\ell}_{\alpha;\tau} &= \inf\left\{t \in [0,1] : \hat{G}_\tau(t) < (1-\alpha)\left(1+\frac{1}{n}\right)\right\} \\
&= \min\left\{x_i : i \in \{1, \ldots, \kappa\}, s_{i-1} < (1-\alpha)\left(1+\frac{1}{n}\right)\right\}.
\end{aligned}$$

□

Proposition 2 (Instant-by-instant prediction upper band, MDist-full)**.** *Let $\xi_1, \ldots, \xi_n$ be n independent copies of ξ. For any control level $\alpha \in [0,1]$, the band $\hat{B}_\alpha^{\mathrm{MDist-full,up}}$ defined as, for every $\tau \in \{1, \ldots, T\}$*

$$\hat{B}_{\alpha;\tau}^{\mathrm{MDist-full,up}} := \left\{k \in \{1, \ldots, K\} : k \leq \hat{Q}_\tau^{\mathrm{up}}\left((1-\alpha)\left(1+\frac{1}{n}\right)\right)+1\right\} \tag{4}$$

ensures control of the instant-by-instant coverage of ξ at a control level α, the empirical marginal upper quantile function $\hat{Q}_\tau^{\mathrm{up}}$ is defined in Equation (Eq. (6)).

Proof of Proposition 2. Let $\alpha \in [0,1]$ and $\tau \in \{1, \ldots, T\}$. A full conformal prediction region $\hat{C}_{\alpha;\tau}$ can be defined as

$$\hat{C}_{\alpha;\tau} := \left\{k \in \{1, \ldots, K\} : \hat{\pi}(k) > \alpha\right\},$$

where for every $k \in \{1, \ldots, K\}$

$$\hat{\pi}(k) := \frac{1 + \sum_{i=1}^{n} \mathbb{1}\left\{\hat{F}_\tau^k(\xi_{i;\tau}) \geq \hat{F}_\tau^k(k)\right\}}{n+1},$$

and for every $l \in \{1, \ldots, K\}$

$$\hat{F}_\tau^k(l) := \frac{n}{n+1}\hat{F}_\tau(l) + \frac{1}{n+1}\mathbb{1}\left\{k \leq l\right\},$$

where the empirical marginal cumulative density function $\hat{F}_\tau$ is defined as

$$\hat{F}_\tau(l) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\xi_{i;\tau} \le l\}. \tag{5}$$

Since $\xi_{1;\tau}, \ldots, \xi_{n;\tau}$ are i.i.d. copies of ξ_τ, $\hat{C}_{\alpha;\tau}$ is a confidence prediction region (Vovk, Gammerman, and Shafer, 2005)

$$\mathbb{P}\left[\xi_\tau \in \hat{C}_{\alpha;\tau}\right] \ge 1-\alpha.$$

Let us now provide an explicit expression of $\hat{C}_{\alpha;\tau}$.

$$\begin{aligned}
k \in \hat{C}_{\alpha;\tau} &\iff \hat{\pi}_\tau(k) > \alpha \\
&\iff \frac{1+\sum_{i=1}^{n}\mathbb{1}\left\{\hat{F}^k_\tau(\xi_{i;\tau}) \ge \hat{F}^k_\tau(k)\right\}}{n+1} > \alpha \\
&\iff \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\left\{\hat{F}^k_\tau(\xi_{i;\tau}) < \hat{F}^k_\tau(k)\right\} < (1-\alpha)\left(1+\frac{1}{n}\right).
\end{aligned}$$

Let us focus on the l.h.s. term. Let us define the function $\hat{G}_\tau : \{1,\ldots,K\} \to [0,1]$ such that for every k

$$\begin{aligned}
\hat{G}_\tau(k) &:= \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\left\{\hat{F}^k_\tau(\xi_{i;\tau}) < \hat{F}^k_\tau(k)\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{K}\mathbb{1}\{l=\xi_{i;\tau}\}\,\mathbb{1}\left\{\hat{F}^k_\tau(\xi_{i;\tau}) < \hat{F}^k_\tau(k)\right\} \\
&= \sum_{l=1}^{K}\hat{p}_\tau(l)\,\mathbb{1}\left\{\hat{F}^k_\tau(l) < \hat{F}^k_\tau(k)\right\} \\
&= \sum_{l=1}^{K}\hat{p}_\tau(l)\,\mathbb{1}\left\{\frac{n}{n+1}\hat{F}_\tau(l) + \frac{1}{n+1}\mathbb{1}\{k \le l\} < \frac{n}{n+1}\hat{F}_\tau(k) + \frac{1}{n+1}\right\} \\
&= \sum_{l=1}^{K}\hat{p}_\tau(l)\,\mathbb{1}\left\{\hat{F}_\tau(l) + \frac{1}{n}\mathbb{1}\{k \le l\} < \hat{F}_\tau(k) + \frac{1}{n}\right\} \\
&= \sum_{l=1, l<k}^{K}\hat{p}_\tau(l)\,\mathbb{1}\left\{\hat{F}_\tau(l) < \hat{F}_\tau(k) + \frac{1}{n}\right\} + \sum_{l=1, l\ge k}^{K}\hat{p}_\tau(l)\,\mathbb{1}\left\{\hat{F}_\tau(l) < \hat{F}_\tau(k)\right\}.
\end{aligned}$$

If $1 \le l < k$ then $\hat{F}_\tau(l) \le \hat{F}_\tau(k) < \hat{F}_\tau(k) + \frac{1}{n}$, and if $K \ge l \ge k$ then $\hat{F}_\tau(l) \ge \hat{F}_\tau(k)$. The above equation can be rewritten as

$$\hat{G}_\tau(k) = \sum_{l=1, l<k}^{K}\hat{p}_\tau(l) = \hat{F}_\tau(k-1).$$

Going back to trying to find an expression for $\hat{C}_{\alpha;\tau}$

$$k \in \hat{C}_{\alpha;\tau} \iff \hat{F}_\tau(k-1) < (1-\alpha)\left(1+\frac{1}{n}\right)$$

$$
\begin{aligned}
&\iff k - 1 \leq \hat{Q}_\tau\left((1-\alpha)\left(1+\frac{1}{n}\right)\right) \\
&\iff k \leq \hat{Q}_\tau\left((1-\alpha)\left(1+\frac{1}{n}\right)\right) + 1,
\end{aligned}
$$

where the empirical marginal upper quantile function is defined as, for every level $a \in [0,1]$,

$$
\hat{Q}^{\mathrm{up}}_\tau(a) := \max\left\{k \in \{1,\dots,K\} : \hat{F}_\tau < a.\right\} \tag{6}
$$

the empirical marginal cumulative distribution function $\hat{F}_\tau$ in Equation (Eq. (5)). $\square$

Proposition 3 (Instant-by-instant prediction lower band, MDist-full)**.** *Let $\xi_1,\dots,\xi_n$ be n independent copies of ξ. For any control level $\alpha \in [0,1]$, the band $\hat{B}^{\mathrm{MDist-full,lo}}_\alpha$ defined as, for every $\tau \in \{1,\dots,T\}$*

$$
\hat{B}^{\mathrm{MDist-full,lo}}_{\alpha;\tau} := \left\{k \in \{1,\dots,K\} : k \geq \hat{Q}^{\mathrm{lo}}_\tau\left(\alpha\left(1+\frac{1}{n}\right)-\frac{1}{n}\right)\right\} \tag{7}
$$

ensures control of the instant-by-instant coverage of ξ at a control level α, where the marginal empirical lower quantile function $\hat{Q}^{\mathrm{lo}}_\tau$ is defined in Equation (Eq. (8)).

Proof of Proposition 3. Let $\alpha \in [0,1]$ and $\tau \in \{1,\dots,T\}$. A full conformal prediction region $\hat{C}_{\alpha;\tau}$ can be defined as

$$
\hat{C}_{\alpha;\tau} := \{k \in \{1,\dots,K\} : \hat{\pi}(k) > \alpha\},
$$

where

$$
\hat{\pi}(k) := \frac{1+\sum_{i=1}^n \mathbb{1}\left\{\hat{F}^k_\tau(\xi_{i;\tau}) \leq \hat{F}^k_\tau(k)\right\}}{n+1}.
$$

Since $\xi_{1;\tau},\dots,\xi_{n;\tau}$ are i.i.d. copies of ξ_τ, $\hat{C}_{\alpha;\tau}$ is a confidence prediction region (Vovk, Gammerman, and Shafer, 2005)

$$
\mathbb{P}\left[\xi_\tau \in \hat{C}_{\alpha;\tau}\right] \geq 1-\alpha.
$$

Let us now provide an explicit expression of $\hat{C}_{\alpha;\tau}$.

$$
\begin{aligned}
k \in \hat{C}_{\alpha;\tau} &\iff \hat{\pi}_\tau(k) > \alpha \\
&\iff \frac{1+\sum_{i=1}^n \mathbb{1}\left\{\hat{F}^k_\tau(\xi_{i;\tau}) \leq \hat{F}^k_\tau(k)\right\}}{n+1} > \alpha \\
&\iff \frac{1}{n}\sum_{i=1}^n \mathbb{1}\left\{\hat{F}^k_\tau(\xi_{i;\tau}) \leq \hat{F}^k_\tau(k)\right\} > \alpha\left(1+\frac{1}{n}\right) - \frac{1}{n}.
\end{aligned}
$$

Let us focus on the l.h.s. term. Let us define the function $\tilde{G}_\tau : \{1,\dots,K\} \to [0,1]$ such that for every k

$$
\begin{aligned}
\tilde{G}_\tau(k) &:= \frac{1}{n}\sum_{i=1}^n \mathbb{1}\left\{\hat{F}^k_\tau(\xi_{i;\tau}) \leq \hat{F}^k_\tau(k)\right\} \\
&= \frac{1}{n}\sum_{i=1}^n \sum_{l=1}^K \mathbb{1}\{l = \xi_{i;\tau}\}\, \mathbb{1}\left\{\hat{F}^k_\tau(\xi_{i;\tau}) \leq \hat{F}^k_\tau(k)\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{l=1}^{K} \hat{p}_\tau(l)\, \mathbb{1}\left\{\hat{F}_\tau^k(l) \le \hat{F}_\tau^k(k)\right\} \\
&= \sum_{l=1}^{K} \hat{p}_\tau(l)\, \mathbb{1}\left\{\frac{n}{n+1}\hat{F}_\tau(l) + \frac{1}{n+1}\mathbb{1}\{k \le l\} \le \frac{n}{n+1}\hat{F}_\tau(k) + \frac{1}{n+1}\right\} \\
&= \sum_{l=1}^{K} \hat{p}_\tau(l)\, \mathbb{1}\left\{\hat{F}_\tau(l) + \frac{1}{n}\mathbb{1}\{k \le l\} \le \hat{F}_\tau(k) + \frac{1}{n}\right\} \\
&= \sum_{l=1, l\le k}^{K} \hat{p}_\tau(l)\, \mathbb{1}\left\{\hat{F}_\tau(l) \le \hat{F}_\tau(k) + \frac{1}{n}\right\} + \sum_{l=1, l>k}^{K} \hat{p}_\tau(l)\, \mathbb{1}\left\{\hat{F}_\tau(l) \le \hat{F}_\tau(k)\right\}.
\end{aligned}
$$

If $1 \le l < k$ then $\hat{F}_\tau(l) \le \hat{F}_\tau(k) < \hat{F}_\tau(k) + \frac{1}{n}$, and if $K \ge l > k$ and $\hat{p}_\tau(l) \ne 0$, then

$$
\hat{F}_\tau(l) = \sum_{\tilde{l}=1}^{l} \hat{p}_\tau\left(\tilde{l}\right) = \sum_{\tilde{l}=1}^{k} \hat{p}_\tau\left(\tilde{l}\right) + \sum_{\tilde{l}=k+1}^{l} \hat{p}_\tau\left(\tilde{l}\right) = \hat{F}_\tau(k) + \sum_{\tilde{l}=k+1}^{l} \hat{p}_\tau\left(\tilde{l}\right) > \hat{F}_\tau(k),
$$

since $\sum_{\tilde{l}=k+1}^{l} \hat{p}_\tau\left(\tilde{l}\right)$ includes $\hat{p}_\tau(l) \ne 0$. Hence, the function $\tilde{G}_\tau$ can be rewritten as

$$
\tilde{G}_\tau(k) = \sum_{l=1, l\le k}^{K} \hat{p}_\tau(l) = \hat{F}_\tau(k).
$$

Going back to trying to find an expression for $\hat{C}_{\alpha;\tau}$

$$
\begin{aligned}
k \in \hat{C}_{\alpha;\tau} &\iff \hat{F}_\tau(k) > \alpha\left(1 + \frac{1}{n}\right) - \frac{1}{n} \\
&\iff k \ge \tilde{Q}_\tau\left(\alpha\left(1 + \frac{1}{n}\right) - \frac{1}{n}\right),
\end{aligned}
$$

where the empirical marginal lower quantile function is defined as, for every level $a \in (0,1)$

$$
\tilde{Q}_\tau^{\mathrm{lo}}(a) := \min\left\{k \in \{1, \ldots, K\} : \hat{F}_\tau(k) > a\right\}, \tag{8}
$$

the empirical marginal cumulative distribution function $\hat{F}_\tau$ in Equation (Eq. (5)). □

Proposition 4 (Instant-by-instant prediction band, MDist-full)**.** *Let $\xi_1, \ldots, \xi_n$ be n independent copies of ξ. For any control level $\alpha \in [0,1]$, and for any $\alpha_{\mathrm{lo}}, \alpha_{\mathrm{up}} \in [0,1]$ such that $\alpha_{\mathrm{lo}} + \alpha_{\mathrm{up}} = \alpha$, the band $\hat{B}_\alpha^{\mathrm{MDist-full}}$ defined as, for every $\tau \in \{1, \ldots, T\}$*

$$
\begin{aligned}
\hat{B}_{\alpha;\tau}^{\mathrm{MDist-full}} := \Bigg\{k \in \{1, \ldots, K\} : \hat{Q}_\tau^{\mathrm{lo}}\left(\alpha_{\mathrm{lo}}\left(1 + \frac{1}{n}\right) - \frac{1}{n}\right) \le k,& \\
k \le \hat{Q}_\tau^{\mathrm{up}}\left((1 - \alpha_{\mathrm{up}})\left(1 + \frac{1}{n}\right)\right) + 1\Bigg\}&,
\end{aligned} \tag{9}
$$

ensures control of the instant-by-instant coverage of ξ at a control level α.

Proof of Proposition 4. Let $\alpha \in [0,1]$ and $\tau \in \{1, \ldots, T\}$. It follows from union bound that

$$
\begin{aligned}
\mathbb{P}\left[\xi_\tau \in \hat{B}_{\alpha;\tau}^{\mathrm{MDist-full}}\right] &= \mathbb{P}\left[\xi_\tau \in \hat{B}_{\alpha_{\mathrm{up}};\tau}^{\mathrm{MDist-full,up}} \cap \hat{B}_{\alpha_{\mathrm{lo}};\tau}^{\mathrm{MDist-full,lo}}\right] \\
&= 1 - \mathbb{P}\left[\xi_\tau \in \left(\hat{B}_{\alpha_{\mathrm{up}};\tau}^{\mathrm{MDist-full,up}}\right)^c \cup \left(\hat{B}_{\alpha_{\mathrm{lo}};\tau}^{\mathrm{MDist-full,lo}}\right)^c\right]
\end{aligned}
$$

$$\geq 1-\left(\mathbb{P}\left[\xi_\tau \in \left(\hat{B}^{\mathrm{MDist-full,up}}_{\alpha_{\mathrm{up}};\tau}\right)^c\right]+\mathbb{P}\left[\xi_\tau \in \left(\hat{B}^{\mathrm{MDist-full,lo}}_{\alpha_{\mathrm{lo}};\tau}\right)^c\right]\right)$$
$$\geq 1-(\alpha_{\mathrm{up}}+\alpha_{\mathrm{lo}}) = 1-\alpha.$$

□

Proposition 5 (Instant-by-instant prediction band, MDist-Split)**.** *Let $\xi_1,\ldots,\xi_n$ be n independent copies of ξ. Let I_1 and I_2 be two index sets with cardinal n_1 and n_2 respectively such that $I_1 \sqcup I_2 = \{1,\ldots,n\}$. For any control level $\alpha\in[0,1]$, the band $\hat{B}^{\mathrm{MDist-split}}_\alpha$ defined as, for every $\tau\in\{1,\ldots,T\}$*

$$\hat{B}^{\mathrm{MDist-split}}_{\alpha;\tau} := \left\{k\in\{1,\ldots,K\} : \tilde{Q}^{\mathrm{lo}}_{D_1;\tau}\left(\hat{t}^{\mathrm{MDist-split}}_{\alpha;\tau}\right) \leq k \leq \tilde{Q}^{\mathrm{up}}_{D_1;\tau}\left(1-\hat{t}^{\mathrm{MDist-split}}_{\alpha;\tau}\right)\right\}, \quad (10)$$

ensures control of the instant-by-instant coverage of ξ at a control level α, where the threshold $\hat{t}^{\mathrm{MDist-split}}_{\alpha;\tau}$ is defined in Equation (Eq. (12)) and the marginal empirical upper and lower quantile function $\tilde{Q}^{\mathrm{up}}_{D_1;\tau}$ and $\tilde{Q}^{\mathrm{lo}}_{D_1;\tau}$ are defined in Equations (Eq. (13)) and (Eq. (14)) respectively.

Proof of Proposition 5. Let $\tau\in\{1,\ldots,T\}$. Define the empirical marginal cumulative density function $\hat{F}_{D_1;\tau}$ trained on the data set $D_1 := \{\xi_i, i\in I_1\}$ as, for every $k\in\{1,\ldots,K\}$

$$\hat{F}_{D_1;\tau}(k) := \frac{1}{n_1}\sum_{i\in I_1}\mathbb{1}\{\xi_{i;\tau}\leq k\}. \quad (11)$$

Upon the dataset $D_2 := \{\xi_i, i\in I_2\}$, one can define split conformal prediction region $\hat{C}_{\alpha;\tau}$ as

$$\hat{C}_{\alpha;\tau} := \{k\in\{1,\ldots,K\} : \hat{\pi}_{D_2,\tau}(k) > \alpha\},$$

where for every $k\in(1,\ldots,K)$

$$\hat{\pi}_{D_2,\tau}(k) := \frac{1+\sum_{i\in \mathrm{I}_2}\mathbb{1}\left\{\hat{A}_{D_1;\tau}(\xi_{i;\tau})\leq \hat{A}_{D_1;\tau}(k)\right\}}{1+n_2},$$

where the conformity score $\hat{A}_{D_1;\tau}(k) := \min\left(\hat{F}_{D_1;\tau}(k), 1-\hat{F}_{D_1;\tau}(k)\right)$. Since $\xi_{1;\tau},\ldots,\xi_{n;\tau}$ are i.i.d. copies of ξ_τ, it follows that $\hat{C}_{\alpha;\tau}$ is a confidence prediction region

$$\mathbb{P}\left(\xi_\tau\in\hat{C}_{\alpha;\tau}\right)\geq 1-\alpha.$$

Let us now compute an explicit expression for $\hat{C}_{\alpha;\tau}$.

$$\begin{aligned}
& k\in\hat{C}_{\alpha;\tau} \\
\Longleftrightarrow & \frac{1+\sum_{i\in \mathrm{I}_2}\mathbb{1}\left\{\hat{A}_{D_1;\tau}(\xi_{i;\tau})\leq \hat{A}_{D_1;\tau}(k)\right\}}{1+n_2} > \alpha \\
\Longleftrightarrow & \frac{1}{n_2}\sum_{i\in \mathrm{I}_2}\mathbb{1}\left\{\hat{A}_{D_1;\tau}(\xi_{i;\tau})\leq \hat{A}_{D_1;\tau}(k)\right\} > \alpha\left(1+\frac{1}{n_2}\right)-\frac{1}{n_2} \\
\Longleftrightarrow & \hat{A}_{D_1;\tau}(k)\geq \hat{\ell}_{\alpha;\tau},
\end{aligned}$$

where the level $\hat{t}^{\mathrm{MDist-split}}_{\alpha;\tau}$ is defined as

$$\hat{t}^{\mathrm{MDist-split}}_{\alpha;\tau} := \inf\left\{t\in[0,1] : \frac{1}{n_2}\sum_{i\in \mathrm{I}_2}\mathbb{1}\left\{\hat{A}_{D_1;\tau}(\xi_{i;\tau})\leq t\right\} > \alpha\left(1+\frac{1}{n_2}\right)-\frac{1}{n_2}\right\}$$

$$= \hat{A}_{D_1;\tau}\left(\xi_{\left(i_{n_2;\alpha}\right),\tau}\right), \tag{12}$$

where the index $i_{n_2;\alpha} := \lceil \alpha (n_2 + 1) - 1 \rceil$, and $\hat{A}_{D_1;\tau}\left(\xi_{(1);\tau}\right) \leq \ldots \leq \hat{A}_{D_1;\tau}\left(\xi_{(n_2);\tau}\right)$ are the conformity scores sorted in increasing order. Going back to trying to find an expression of the prediction region

$$\begin{aligned}
&k \in \hat{C}_{\alpha;\tau} \\
\Longleftrightarrow &\min\left(\hat{F}_{D_1;\tau}(k), 1 - \hat{F}_{D_1;\tau}(k)\right) \geq \hat{t}_{\alpha;\tau}^{\mathrm{MDist-split}}, \\
\Longleftrightarrow &\hat{F}_{D_1;\tau}(k) \geq \hat{t}_{\alpha;\tau}^{\mathrm{MDist-split}}, \text{ and } 1 - \hat{F}_{D_1;\tau}(k) \geq \hat{t}_{\alpha;\tau}^{\mathrm{MDist-split}}, \\
\Longleftrightarrow &\hat{t}_{\alpha;\tau}^{\mathrm{MDist-split}} \leq \hat{F}_{D_1;\tau}(k) \leq 1 - \hat{t}_{\alpha;\tau}^{\mathrm{MDist-split}}, \\
\Longleftrightarrow &\tilde{Q}^{\mathrm{lo}}_{D_1;\tau}\left(\hat{t}_{\alpha;\tau}^{\mathrm{MDist-split}}\right) \leq k \leq \tilde{Q}^{\mathrm{up}}_{D_1;\tau}\left(1 - \hat{t}_{\alpha;\tau}^{\mathrm{MDist-split}}\right),
\end{aligned}$$

where the empirical marginal upper quantile function built from D_1 is defined as, for every level $a \in [0, 1]$ is defined as

$$\tilde{Q}^{\mathrm{up}}_{D_1;\tau}(1 - a) := \max\left\{k \in \{1, \ldots, K\} : \hat{F}_{D_1;\tau}(k) \leq 1 - a\right\}, \tag{13}$$

as for the lower one

$$\tilde{Q}^{\mathrm{lo}}_{D_1;\tau}(a) := \min\left\{k \in \{1, \ldots, K\} : \hat{F}_{D_1;\tau}(k) \geq a\right\}. \tag{14}$$

□

Proposition 6 (γ-simultaneous band)**.** *Let $\xi_1, \ldots, \xi_n$ be i.i.d. copies of ξ. Let I_1 and I_2 be two index sets with cardinal n_1 and n_2 respectively such that $I_1 \sqcup I_2 = \{1, \ldots, n\}$. Let $\hat{A}_{D_1;\tau} : \{1, \ldots, K\} \to \mathbb{R}$, $k \mapsto \hat{A}_{D_1;\tau}(k)$ be a conformity-measure over $\{1, \ldots, K\}$ built from the dataset $D_1 := \{\xi_i : i \in I_1\}$.*

For any slack level $\gamma \in [0, 1]$, control level $\alpha \in [0, 1]$, the prediction band $\hat{B}_{\gamma;\alpha}$ defined as for every $\tau \in \{1, \ldots, T\}$

$$\hat{B}_{\gamma;\alpha;\tau} := \left\{k \in \{1, \ldots, K\} : \hat{A}_{D_1;\tau}(k) \geq \hat{t}_{D_1;\gamma}\left(\xi_{\left(i_{n_2;\alpha}\right)}\right)\right\}$$

ensures control of the γ-simultaneous coverage of ξ at a level α where the cut-off level $\hat{t}_{D_1;\gamma}\left(\xi_{\left(i_{n_2;\alpha}\right)}\right)$ is defined in Equation (Eq. 16) from the conformity score defined in Equation (Eq. 15).

Proof of Proposition 6. Let $\gamma \in [0, 1]$ and $\alpha \in [0, 1]$. One can define a conformity measure $\hat{t}_{D_1;\gamma} : \{1, \ldots, K\}^T \to \mathbb{R}$ such that for every $\xi \in \{1, \ldots, K\}^T$

$$\hat{t}_{D_1;\gamma}(\xi) := \sup\left\{t \in \mathbb{R} : \frac{1}{T}\sum_{\tau=1}^{T} \mathbb{1}\left\{\hat{A}_{D_1;\tau}(\xi_\tau) \geq t\right\} \geq 1 - \gamma\right\} = \hat{A}_{D_1;\left(\tau_{\gamma,T}\right)}\left(\xi_{\left(\tau_{\gamma,T}\right)}\right), \tag{15}$$

where $\hat{A}_{D_1;(1)}\left(\xi_{(1)}\right) \geq \ldots \geq \hat{A}_{D_1;(T)}\left(\xi_{(T)}\right)$ are sorted in decreasing order and $\tau_{\gamma,T} := \lceil T(1 - \gamma) \rceil$. Upon the dataset $D_2 := \{\xi_i : i \in I_2\}$, one can define a split conformal prediction region as

$$\hat{C}_{\gamma;\alpha} := \left\{\xi \in \{1, \ldots, K\}^T : \hat{\pi}_{D_2}(\xi) > \alpha\right\},$$

where for every $\xi \in \{1, \ldots, K\}^T$

$$\hat{\pi}_{D_2}(\xi) := \frac{1 + \sum_{i \in I_2}\left\{\hat{t}_{D_1;\gamma}(\xi_i) \leq \hat{t}_{D_1;\gamma}(\xi)\right\}}{n_2 + 1}.$$

Let us now provide an explicit expression for $\hat{C}_{\gamma;\alpha}$,

$$
\begin{aligned}
\xi \in \hat{C}_{\gamma;\alpha} &\iff \frac{1+\sum_{i\in I_2}\left\{\hat{t}_{D_1;\gamma}(\xi_i) \le \hat{t}_{D_1;\gamma}(\xi)\right\}}{n_2+1} > \alpha \\
&\iff \frac{1}{n_2}\sum_{i\in I_2}\left\{\hat{t}_{D_1;\gamma}(\xi_i) \le \hat{t}_{D_1;\gamma}(\xi)\right\} > \left(1+\frac{1}{n_2}\right)\alpha - \frac{1}{n_2} \\
&\iff \hat{t}_{D_1;\gamma}(\xi) \ge \hat{t}_{D_1;\gamma}\left(\xi_{\left(i_{n_2;\alpha}\right)}\right), \qquad (16)
\end{aligned}
$$

where $\hat{t}_{D_1;\gamma}\left(\xi_{\left(i_{n_2;\alpha}\right)}\right) \le \ldots \le \hat{t}_{D_1;\gamma}\left(\xi_{\left(i_{n_2;\alpha}\right)}\right)$ are sorted in increasing order and $i_{n_2;\alpha} := \lceil (n_2+1)\alpha - 1 \rceil$. Hence, the prediction band $\hat{B}_{\gamma;\alpha}$ defined as for every $\tau \in \{1,\ldots,T\}$

$$
\hat{B}_{\gamma;\alpha;\tau} := \left\{k \in \{1,\ldots,K\} : \hat{A}_{D_1;\tau}(k) \ge \hat{t}_{D_1;\gamma}\left(\xi_{\left(i_{n_2;\alpha}\right)}\right)\right\}
$$

ensures the following

$$
\begin{aligned}
\mathbb{P}\left[\frac{1}{T}\sum_{\tau=1}^{T}\mathbb{1}\left\{\xi_\tau \in \hat{B}_{\gamma;\alpha;\tau}\right\}\right] &= \mathbb{P}\left[\frac{1}{T}\sum_{\tau=1}^{T}\mathbb{1}\left\{\hat{A}_{D_1;\tau}(\xi_\tau) \ge \hat{t}_{D_1;\gamma}\left(\xi_{\left(i_{n_2;\alpha}\right)}\right)\right\}\right] \\
&= \mathbb{P}\left[\hat{t}_{D_1;\gamma}(\xi) \ge \hat{t}_{D_1;\gamma}\left(\xi_{\left(i_{n_2;\alpha}\right)}\right)\right] \\
&= \mathbb{P}\left[\xi \in \hat{C}_{\gamma,\alpha}\right] \ge 1-\alpha.
\end{aligned}
$$

□

Corollary 1 (γ-simultaneous band, MD-Split)**.** *Let $\xi_1,\ldots,\xi_n$ be i.i.d. copies of ξ. Let I_1 and I_2 be two index sets with cardinal n_1 and n_2 respectively such that $I_1 \sqcup I_2 = \{1,\ldots,n\}$.*

For any slack $\gamma \in (0,1)$, control level $\alpha \in [0,1]$, the prediction band $\hat{B}^{\mathrm{MD-split}}_{\gamma;\alpha}$ such that for every $\tau \in \{1,\ldots,T\}$

$$
\hat{B}^{\mathrm{MD-split}}_{\gamma;\alpha;\tau} := \left\{k \in \{1,\ldots,K\} : \hat{p}_{D_1;\tau}(k) \ge \hat{\ell}^{\mathrm{MD-Split}}_{\gamma;\alpha}\right\}, \qquad (17)
$$

ensures control of γ-simultaneous coverage of ξ at a control level α, where empirical marginal probability density function $\hat{p}_{D_1;\tau}$ is defined in Equation (Eq. (18)), and the threshold is $\hat{\ell}^{\mathrm{MD-Split}}_{\gamma;\alpha}$ is defined in Equation (Eq. (19)).

Proof of Corollary 1. For every $\tau \in \{1,\ldots,K\}$, define empirical marginal probability density function $\hat{p}_{D_1;\tau}$ built from the dataset $D_1 := \{\xi_i : i \in I_1\}$ as, for every $k \in \{1,\ldots,K\}$

$$
\hat{p}_{D_1;\tau}(k) := \frac{1}{n_1}\sum_{i\in I_1}\mathbb{1}\{\xi_{i;\tau} = k\}. \qquad (18)
$$

Applying Proposition 6 by choosing for every $\tau \in \{1,\ldots,T\}$, and every $k \in \{1,\ldots,K\}$, $\hat{A}_{D_1;\tau}(k) := \hat{p}_{D_1;\tau}(k)$, the prediction band $\hat{B}^{\mathrm{MD-split}}_{\gamma;\alpha}$ defined as, for every $\tau \in \{1,\ldots,T\}$

$$
\begin{aligned}
\hat{B}^{\mathrm{MD-split}}_{\gamma;\alpha;\tau} &:= \left\{k \in \{1,\ldots,K\} : \hat{A}_{D_1;\tau}(k) \ge \hat{t}_{D_1;\gamma}\left(\xi_{\left(i_{n_2;\alpha}\right)}\right)\right\} \\
&= \left\{k \in \{1,\ldots,K\} : \hat{p}_{D_1;\tau}(k) \ge \hat{\ell}^{\mathrm{MD-Split}}_{\gamma;\alpha}\right\},
\end{aligned}
$$

ensures control of γ-simultaneous coverage of ξ at a control level α, where the threshold $\hat{\ell}^{\mathrm{MD-Split}}_{\gamma;\alpha}$ is defined as

$$
\hat{\ell}^{\mathrm{MD-Split}}_{\gamma;\alpha} := \hat{t}_{D_1;\gamma}\left(\xi_{\left(i_{n_2;\alpha}\right)}\right). \qquad (19)
$$

□

Corollary 2 (γ-simultaneous band, MHPD-Split)**.** *Let $\xi_1, \ldots, \xi_n$ be i.i.d. copies of ξ. Let I_1 and I_2 be two index sets with cardinal n_1 and n_2 respectively such that $I_1 \sqcup I_2 = \{1, \ldots, n\}$.*

For any slack $\gamma \in (0, 1)$, control level $\alpha \in [0, 1]$, the prediction band $\hat{B}^{\mathrm{MHPD-split}}_{\gamma;\alpha}$ such that for every $\tau \in \{1, \ldots, T\}$

$$\hat{B}^{\mathrm{MHPD-split}}_{\gamma;\alpha;\tau} := \left\{k \in \{1, \ldots, K\} : \hat{p}_{D_1;\tau}(k) > \hat{\ell}^{\mathrm{MHPD-split}}_{\gamma;\alpha;\tau}\right\}, \tag{20}$$

ensures γ-simultaneous coverage of ξ at a control level α, where empirical marginal probability density function $\hat{p}_{D_1;\tau}$ is defined in Equation (Eq. (18)), and the threshold $\hat{\ell}^{\mathrm{MHPD-split}}_{\gamma;\alpha;\tau}$ is defined in Equation (Eq. (22)).

Proof of Corollary 2. For every $\tau \in \{1, \ldots, K\}$, and every $k \in \{1, \ldots, K\}$ define the rank $\hat{R}_{D_1;\tau}(k)$ of k at the instant τ built from the dataset $D_1 := \{\xi_i : i \in I_1\}$ as

$$\hat{R}_{D_1;\tau}(k) := \sum_{l=1}^{K} \hat{p}_{D_1;\tau}(l)\, \mathbb{1}\left\{\hat{p}_{D_1;\tau}(l) \leq \hat{p}_{D_1;\tau}(k)\right\}. \tag{21}$$

Applying Proposition 6 by choosing for every $\tau \in \{1, \ldots, T\}$, and every $k \in \{1, \ldots, K\}$, $\hat{A}_{D_1;\tau}(k) := \hat{R}_{D_1;\tau}(k)$, the prediction band $\hat{B}^{\mathrm{MD-split}}_{\gamma;\alpha}$ defined as, for every $\tau \in \{1, \ldots, T\}$

$$\begin{aligned}
\hat{B}^{\mathrm{MHPD-split}}_{\gamma;\alpha;\tau} &:= \left\{k \in \{1, \ldots, K\} : \hat{A}_{D_1;\tau}(k) \geq \hat{t}_{D_1;\gamma}\left(\xi_{(i_{n_2;\alpha})}\right)\right\} \\
&= \left\{k \in \{1, \ldots, K\} : \hat{R}_{D_1;\tau}(k) \geq \hat{t}_{D_1;\gamma}\left(\xi_{(i_{n_2;\alpha})}\right)\right\} \\
&= \left\{k \in \{1, \ldots, K\} : \hat{p}_{D_1;\tau}(k) > \hat{\ell}^{\mathrm{MHPD-split}}_{\gamma;\alpha;\tau}\right\},
\end{aligned}$$

ensures control of γ-simultaneous coverage of ξ at a control level α, where the cut-off $\hat{\ell}^{\mathrm{MHPD-split}}_{\gamma;\alpha;\tau}$ is defined as

$$\hat{\ell}^{\mathrm{MHPD-split}}_{\gamma;\alpha;\tau} := \sup\left\{\ell \in [0, 1] : \sum_{l=1}^{K} \hat{p}_{D_1;\tau}(l)\, \mathbb{1}\left\{\hat{p}_{D_1;\tau}(l) > \ell\right\} > 1 - \hat{t}^{\mathrm{MHPD-split}}_{D_1;\gamma}\left(\xi_{(i_{n_2;\alpha})}\right)\right\}, \tag{22}$$

□

Corollary 3 (γ-simultaneous band, MDist-Split)**.** *Let $\xi_1, \ldots, \xi_n$ be n independent copies of ξ. Let I_1 and I_2 be two index sets with cardinal n_1 and n_2 respectively such that $I_1 \sqcup I_2 = \{1, \ldots, n\}$.*

For any slack $\gamma \in (0, 1)$, any control level $\alpha \in [0, 1]$, the band $\hat{B}^{\mathrm{MDist-split}}_{\gamma;\alpha}$ defined as, for every $\tau \in \{1, \ldots, T\}$

$$\hat{B}^{\mathrm{MDist-split}}_{\gamma;\alpha;\tau} := \left\{k \in \{1, \ldots, K\} : \hat{Q}^{\mathrm{lo}}_{D_1,\tau}\left(\hat{t}^{\mathrm{MDist-split}}_{\gamma;\alpha}\right) \leq k \leq \hat{Q}^{\mathrm{lo}}_{D_1,\tau}\left(1 - \hat{t}^{\mathrm{MDist-split}}_{\gamma;\alpha}\right)\right\}, \tag{23}$$

ensures γ-simultaneous coverage of ξ at a control level α, where the level $\hat{t}^{\mathrm{MDist-split}}_{\gamma;\alpha}$ is defined in Equation (Eq. (25)).

Proof of Corollary 3. Applying Proposition 6, for every $\tau \in \{1, \ldots, K\}$, and $k \in \{1, \ldots, K\}$

$$\hat{A}_{D_1;\tau}(k) := \min\left(\hat{F}_{D_1;\tau}(k), 1 - \hat{F}_{D_1;\tau}(k)\right), \tag{24}$$

the prediction band $\hat{B}^{\mathrm{MDist-split}}_{\gamma;\alpha}$ defined as, for every $\tau \in \{1, \ldots, T\}$

$$\hat{B}^{\mathrm{MHPD-split}}_{\gamma;\alpha;\tau} := \left\{k \in \{1, \ldots, K\} : \hat{A}_{D_1;\tau}(k) \geq \hat{t}_{D_1;\gamma}\left(\xi_{(i_{n_2;\alpha})}\right)\right\}$$

$$= \left\{ k \in \{1, \ldots, K\} : \min \left(\hat{F}_{D_1;\tau}(k), 1 - \hat{F}_{D_1;\tau}(k) \right) \geq \hat{t}_{\gamma;\alpha}^{\mathrm{MDist-split}} \right\}$$
$$= \left\{ k \in \{1, \ldots, K\} : \hat{Q}_{D_1,\tau}^{\mathrm{lo}} \left(\hat{t}_{\gamma;\alpha}^{\mathrm{MDist-split}} \right) \leq k \leq \hat{Q}_{D_1,\tau}^{\mathrm{lo}} \left(1 - \hat{t}_{\gamma;\alpha}^{\mathrm{MDist-split}} \right) \right\},$$

ensures control of γ-simultaneous coverage of ξ at a control level α, where the level $\hat{t}_{\gamma;\alpha}^{\mathrm{MDist-split}}$ is defined as

$$\hat{t}_{\gamma;\alpha}^{\mathrm{MDist-split}} := \hat{t}_{D_1;\gamma} \left(\xi_{\left(i_{n_2;\alpha}\right)} \right). \tag{25}$$

□

References

[1] Jing Lei, James Robins, and Larry Wasserman. "Distribution-free prediction sets". In: *Journal of the American Statistical Association* 108.501 (2013), pp. 278–287.

[2] Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*. Vol. 29. Springer, 2005.