Universal distribution of the coverage in split conformal prediction

Paulo C. Marques F.

<PauloCMF1@insper.edu.br>

December 14, 2023

Insper

V. Vovk, A. Gammerman, and C. Saunders, "Machine-learning applications of algorithmic randomness," in *Proceedings of the Sixteenth International Conference on Machine Learning*, ICML '99, (San Francisco, CA, USA), pp. 444–453, Morgan Kaufmann Publishers Inc., 1999.

H. Papadopoulos, K. Proedrou, V. Vovk, and A. Gammerman, "Inductive confidence machines for regression," in *Machine Learning: ECML 2002* (T. Elomaa, H. Mannila, and H. Toivonen, eds.), (Berlin, Heidelberg), pp. 345–356, Springer Berlin Heidelberg, 2002.

V. Vovk, A. Gammerman, and G. Shafer, *Algorithmic learning in a random world*. Springer Science & Business Media, 2005.

J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman, "Distribution-free predictive inference for regression," *Journal of the American Statistical Association*, vol. 113, no. 523, pp. 1094–1111, 2018.

Initial intuitive discussion

- Here we consider **Split** Conformal Prediction

Initial intuitive discussion

Insper

- Here we consider **Split** Conformal Prediction
- Given a predictive model: $\hat{\psi} : \mathbb{R}^d \to \mathbb{R} : (x_{i1}, \ldots, x_{id}) \mapsto y_i$

Initial intuitive discussion

- Here we consider **Split** Conformal Prediction
- Given a predictive model: $\hat{\psi} : \mathbb{R}^d \to \mathbb{R} : (x_{i1}, \ldots, x_{id}) \mapsto y_i$
- Random sample: $(X_1, Y_1), \ldots, (X_n, Y_n), (X_{n+1}, Y_{n+1})$

Initial intuitive discussion

- Here we consider **Split** Conformal Prediction
- Given a predictive model: $\hat{\psi} : \mathbb{R}^d \to \mathbb{R} : (x_{i1}, \ldots, x_{id}) \mapsto y_i$
- Random sample: $(X_1, Y_1), \ldots, (X_n, Y_n), (X_{n+1}, Y_{n+1})$
- Conformity scores: $R_i = |Y_i - \hat{\psi}(X_i)|$

Initial intuitive discussion

- Here we consider **Split** Conformal Prediction
- Given a predictive model: $\hat{\psi} : \mathbb{R}^d \to \mathbb{R} : (x_{i1}, \ldots, x_{id}) \mapsto y_i$
- Random sample: $(X_1, Y_1), \ldots, (X_n, Y_n), (X_{n+1}, Y_{n+1})$
- Conformity scores: $R_i = |Y_i - \hat{\psi}(X_i)|$
- Ordered **calibration** scores:

$$R_{n+1}$$

$$\_\_\_\_\_ \quad R_{(1)} \quad < \quad R_{(2)} \quad < \quad \cdots \quad < \quad R_{(n)} \quad \_\_\_\_\_$$

Initial intuitive discussion

Insper

- Here we consider **Split** Conformal Prediction
- Given a predictive model: $\hat{\psi} : \mathbb{R}^d \to \mathbb{R} : (x_{i1}, \ldots, x_{id}) \mapsto y_i$
- Random sample: $(X_1, Y_1), \ldots, (X_n, Y_n), (X_{n+1}, Y_{n+1})$
- Conformity scores: $R_i = |Y_i - \hat{\psi}(X_i)|$
- Ordered **calibration** scores:

$$R_{n+1}$$

$$\_\_\_\_\_\_ \quad R_{(1)} \quad < \quad R_{(2)} \quad < \quad \cdots \quad < \quad R_{(n)} \quad \_\_\_\_\_\_$$

$$P\left(R_{n+1} < R_{(k)}\right) = \frac{k}{n+1} \qquad\qquad k = 1, \ldots, n$$

Initial intuitive discussion

- Nominal “miscoverage” level: $0 < \alpha < 1$

Initial intuitive discussion

Insper

- Nominal "miscoverage" level: $0 < \alpha < 1$

$$(1-\alpha)(n+1) \leq \lceil (1-\alpha)(n+1) \rceil \leq (1-\alpha)(n+1) + 1$$

Initial intuitive discussion

Insper

- Nominal "miscoverage" level: $0 < \alpha < 1$

$$(1-\alpha)(n+1) \leq \lceil (1-\alpha)(n+1) \rceil \leq (1-\alpha)(n+1) + 1$$

- Choosing $k = \lceil (1-\alpha)(n+1) \rceil$:

$$1 - \alpha \leq P\left(R_{n+1} < R_{(\lceil (1-\alpha)(n+1) \rceil)}\right) \leq 1 - \alpha + \frac{1}{n+1}$$

Initial intuitive discussion

Insper

- Nominal "miscoverage" level: $0 < \alpha < 1$

$$(1-\alpha)(n+1) \leq \lceil (1-\alpha)(n+1) \rceil \leq (1-\alpha)(n+1) + 1$$

- Choosing $k = \lceil (1-\alpha)(n+1) \rceil$:

$$1 - \alpha \leq P\left(R_{n+1} < R_{(\lceil (1-\alpha)(n+1) \rceil)}\right) \leq 1 - \alpha + \frac{1}{n+1}$$

$$\hat{s} = R_{(\lceil (1-\alpha)(n+1) \rceil)}$$

Initial intuitive discussion

Insper

- Nominal "miscoverage" level: $0 < \alpha < 1$

$$(1-\alpha)(n+1) \leq \lceil (1-\alpha)(n+1) \rceil \leq (1-\alpha)(n+1) + 1$$

- Choosing $k = \lceil (1-\alpha)(n+1) \rceil$:

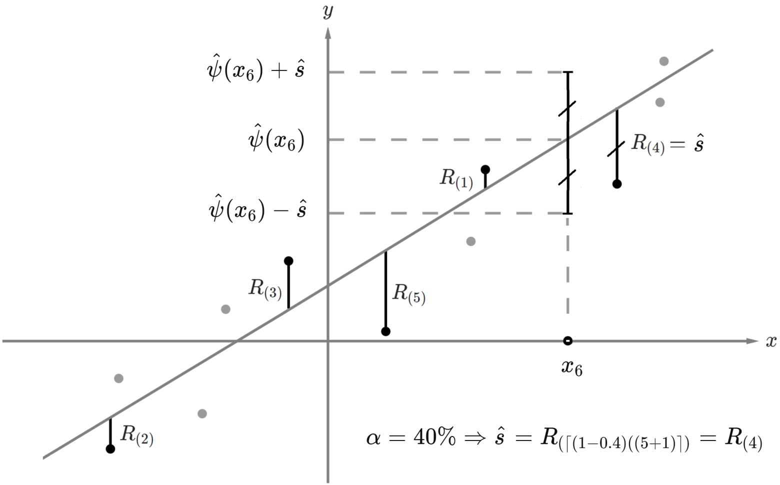
$$1 - \alpha \leq P\left(R_{n+1} < R_{(\lceil (1-\alpha)(n+1) \rceil)}\right) \leq 1 - \alpha + \frac{1}{n+1}$$

$$\hat{s} = R_{(\lceil (1-\alpha)(n+1) \rceil)}$$

- **Marginal Validity Property (MVP):**

$$\boxed{1 - \alpha \leq P\left(\hat{\psi}(X_{n+1}) - \hat{s} < Y_{n+1} < \hat{\psi}(X_{n+1}) + \hat{s}\right) \leq 1 - \alpha + \frac{1}{n+1}}$$

Insper



y
$\hat{\psi}(x_6) + \hat{s}$
$\hat{\psi}(x_6)$
$\hat{\psi}(x_6) - \hat{s}$
$R_{(1)}$
$R_{(4)} = \hat{s}$
$R_{(3)}$
$R_{(5)}$
x
x_6
$R_{(2)}$
$\alpha = 40\% \Rightarrow \hat{s} = R_{(\lceil(1-0.4)((5+1)\rceil)} = R_{(4)}$

What does the MVP imply?

- Let's run a simulation.

What does the MVP imply?

- Let's run a simulation.

- Data Generating Process (DGP): we have ten independent predictors $X_{i,1}, \ldots, X_{i,10}$ with U$[0,1]$ distribution, and an independent random variable ϵ_i with standard normal distribution. The response variable is defined as

$$Y_i = 10\sin(\pi X_{i,1}X_{i,2}) + 20(X_{i,3} - 1/2)^2 + 10X_{i,4} + 5X_{i,5} + \epsilon_i,$$

  implying that the response is not related to the last five predictors, which act as noise in the data.

What does the MVP imply?

Insper

- Let's run a simulation.

- Data Generating Process (DGP): we have ten independent predictors $X_{i,1}, \ldots, X_{i,10}$ with U$[0,1]$ distribution, and an independent random variable ϵ_i with standard normal distribution. The response variable is defined as

$$Y_i = 10\sin(\pi X_{i,1}X_{i,2}) + 20(X_{i,3} - 1/2)^2 + 10X_{i,4} + 5X_{i,5} + \epsilon_i,$$

  implying that the response is not related to the last five predictors, which act as noise in the data.

- We run 10 000 independent replications of this DGP.

What does the MVP imply?

- Let's run a simulation.

- Data Generating Process (DGP): we have ten independent predictors $X_{i,1}, \dots, X_{i,10}$ with U$[0,1]$ distribution, and an independent random variable ϵ_i with standard normal distribution. The response variable is defined as

$$Y_i = 10 \sin(\pi X_{i,1} X_{i,2}) + 20(X_{i,3} - 1/2)^2 + 10 X_{i,4} + 5 X_{i,5} + \epsilon_i,$$

  implying that the response is not related to the last five predictors, which act as noise in the data.

- We run 10 000 independent replications of this DGP.

- A Random Forest of 500 trees is used as our regression model. For each independent replication, we use a training sample of size 100, a calibration sample of size $n = 10$, and a nominal coverage level $1 - \alpha = 0.8$.

What does the MVP imply?

- In this case, the probability of Y_{n+1} being inside the prediction interval is bounded by the MVP to be between 80% and 89.1%.

What does the MVP imply?

- In this case, the probability of Y_{n+1} being inside the prediction interval is bounded by the MVP to be between 80% and 89.1%.

- In the simulation, we observe Y_{n+1} inside the prediction interval in 81.7% of the 10 000 replications.

What does the MVP imply?

Insper

- In this case, the probability of Y_{n+1} being inside the prediction interval is bounded by the MVP to be between 80% and 89.1%.
- In the simulation, we observe Y_{n+1} inside the prediction interval in 81.7% of the 10 000 replications.
- Now, suppose that we extend our future horizon to m predictions.

$$(X_1, Y_1), \ldots, (X_n, Y_n), (X_{n+1}, Y_{n+1}), \ldots, (X_{n+m}, Y_{n+m})$$

What does the MVP imply?

Insper

- In this case, the probability of Y_{n+1} being inside the prediction interval is bounded by the MVP to be between 80% and 89.1%.

- In the simulation, we observe Y_{n+1} inside the prediction interval in 81.7% of the 10 000 replications.

- Now, suppose that we extend our future horizon to m predictions.

$$(X_1, Y_1), \dots, (X_n, Y_n), (X_{n+1}, Y_{n+1}), \dots, (X_{n+m}, Y_{n+m})$$

- Define the indicators $Z_i = 1$, if Y_{n+i} belongs to the corresponding prediction set, and and $Z_i = 0$, otherwise.

What does the MVP imply?

Insper

- In this case, the probability of Y_{n+1} being inside the prediction interval is bounded by the MVP to be between 80% and 89.1%.

- In the simulation, we observe Y_{n+1} inside the prediction interval in 81.7% of the 10 000 replications.

- Now, suppose that we extend our future horizon to m predictions.

$$(X_1, Y_1), \ldots, (X_n, Y_n), (X_{n+1}, Y_{n+1}), \ldots, (X_{n+m}, Y_{n+m})$$

- Define the indicators $Z_i = 1$, if Y_{n+i} belongs to the corresponding prediction set, and and $Z_i = 0$, otherwise.

- Define the **future coverage** associated with this horizon as

$$C_m^{(n,\alpha)} = \frac{1}{m} \sum_{i=1}^{m} Z_i$$

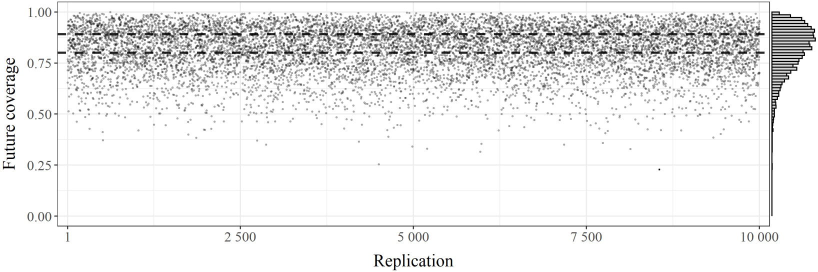
What does the MVP imply?

- What can we know about the distribution of the future coverage $C_m^{(n,\alpha)}$? How is it related to the MVP?

What does the MVP imply?

Insper

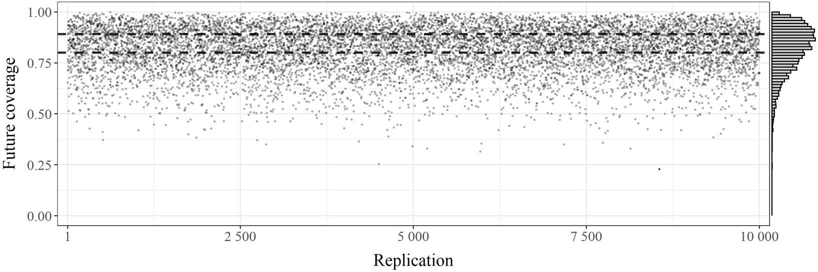
- What can we know about the distribution of the future coverage $C_m^{(n,\alpha)}$? How is it related to the MVP?



What does the MVP imply?

Insper

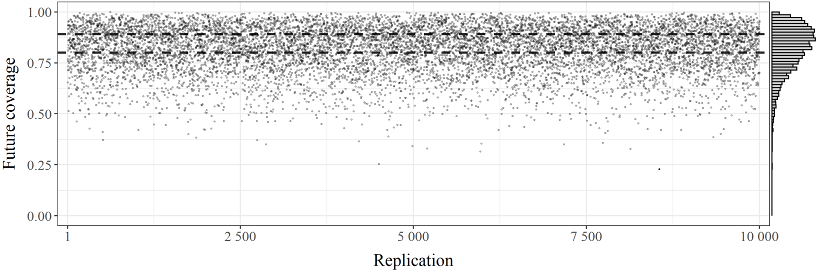
- What can we know about the distribution of the future coverage $C_m^{(n,\alpha)}$? How is it related to the MVP?



- In this simulation we have a horizon of $m = 1\,000$ observations. On replication $8\,550$ we observe the lowest future coverage value: 22.9%.

What does the MVP imply?

- What can we know about the distribution of the future coverage $C_m^{(n,\alpha)}$? How is it related to the MVP?



- In this simulation we have a horizon of $m = 1\,000$ observations. On replication 8 550 we observe the lowest future coverage value: 22.9%.
- The black dashed lines indicate the lower (80%) and upper (89.1%) bounds appearing in the MVP.

What does the MVP imply?

- The MVP has been the most emphasized property of split conformal prediction.

What does the MVP imply?

Insper

- The MVP has been the most emphasized property of split conformal prediction.
- Vovk (2013) discussed the distribution of the future coverage.

What does the MVP imply?

Insper

- The MVP has been the most emphasized property of split conformal prediction.
- Vovk (2013) discussed the distribution of the future coverage.
- An understanding of the distribution of the future coverage $C_m^{(n,\alpha)}$ is important to determine the necessary calibration sample size in applications.

What does the MVP imply?

Insper

- The MVP has been the most emphasized property of split conformal prediction.
- Vovk (2013) discussed the distribution of the future coverage.
- An understanding of the distribution of the future coverage $C_m^{(n,\alpha)}$ is important to determine the necessary calibration sample size in applications.
- *The main contribution of our paper is to make clear the role of exchangeability properties on the determination of the exact distribution of the future coverage $C_m^{(n,\alpha)}$ and the exact distribution of its almost sure limit when the horizon length m grows unboundedly.*
- https://arxiv.org/abs/2303.02770

Exchangeability to the end

Definition. A sequence of random objects $\{O_i\}_{i \geq 1}$ is *exchangeable* if, for every $n \geq 1$, and every permutation $\pi : \{1, \ldots, n\} \xrightarrow{\cong} \{1, \ldots, n\}$, the random tuples $(O_1, \ldots, O_n)$ and $(O_{\pi(i)}, \ldots, O_{\pi(n)})$ have the same distribution.

Exchangeability to the end

Insper

Definition. A sequence of random objects $\{O_i\}_{i \geq 1}$ is *exchangeable* if, for every $n \geq 1$, and every permutation $\pi : \{1, \ldots, n\} \xrightarrow{\cong} \{1, \ldots, n\}$, the random tuples $(O_1, \ldots, O_n)$ and $(O_{\pi(i)}, \ldots, O_{\pi(n)})$ have the same distribution.

- IID $\Rightarrow$ Exchangeable, but the converse is false

Exchangeability to the end

Insper

Definition. A sequence of random objects $\{O_i\}_{i\geq 1}$ is *exchangeable* if, for every $n \geq 1$, and every permutation $\pi : \{1, \dots, n\} \xrightarrow{\cong} \{1, \dots, n\}$, the random tuples $(O_1, \dots, O_n)$ and $(O_{\pi(i)}, \dots, O_{\pi(n)})$ have the same distribution.

- IID $\Rightarrow$ Exchangeable, but the converse is false

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}\right)$$

$$(U_1, U_2) \sim (U_2, U_1) \qquad \text{but} \qquad U_1 \not\perp\!\!\!\perp U_2$$

Exchangeability to the end

Insper

Definition. A sequence of random objects $\{O_i\}_{i\geq 1}$ is *exchangeable* if, for every $n \geq 1$, and every permutation $\pi : \{1,\ldots,n\} \xrightarrow{\cong} \{1,\ldots,n\}$, the random tuples $(O_1,\ldots,O_n)$ and $(O_{\pi(i)},\ldots,O_{\pi(n)})$ have the same distribution.

- IID $\Rightarrow$ Exchangeable, but the converse is false

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \right)$$

$$(U_1, U_2) \sim (U_2, U_1) \qquad \text{but} \qquad U_1 \not\perp\!\!\!\perp U_2$$

- Exchangeable $\Rightarrow$ Identically Distributed, since $(U_i, U_j) \sim (U_j, U_i)$ entails $U_i \sim U_j$

Exchangeability to the end

Insper

Definition. A sequence of random objects $\{O_i\}_{i\geq 1}$ is *exchangeable* if, for every $n \geq 1$, and every permutation $\pi : \{1, \ldots, n\} \xrightarrow{\cong} \{1, \ldots, n\}$, the random tuples $(O_1, \ldots, O_n)$ and $(O_{\pi(i)}, \ldots, O_{\pi(n)})$ have the same distribution.

- IID $\Rightarrow$ Exchangeable, but the converse is false

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}\right)$$

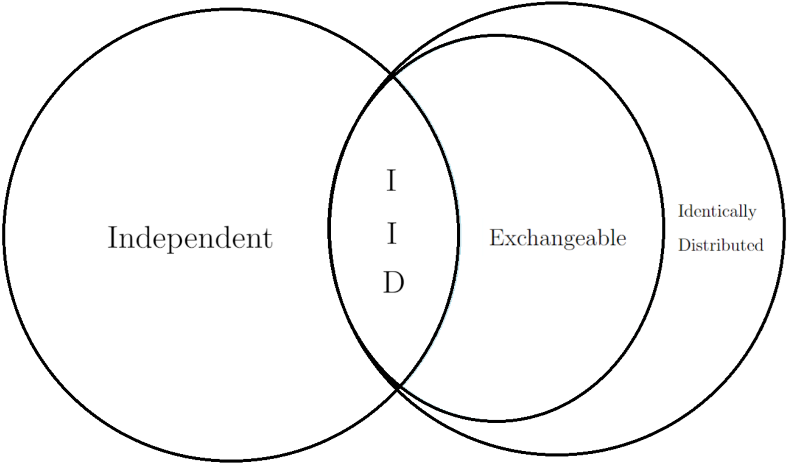
$$(U_1, U_2) \sim (U_2, U_1) \qquad \text{but} \qquad U_1 \not\perp\!\!\!\perp U_2$$

- Exchangeable $\Rightarrow$ Identically Distributed, since $(U_i, U_j) \sim (U_j, U_i)$ entails $U_i \sim U_j$

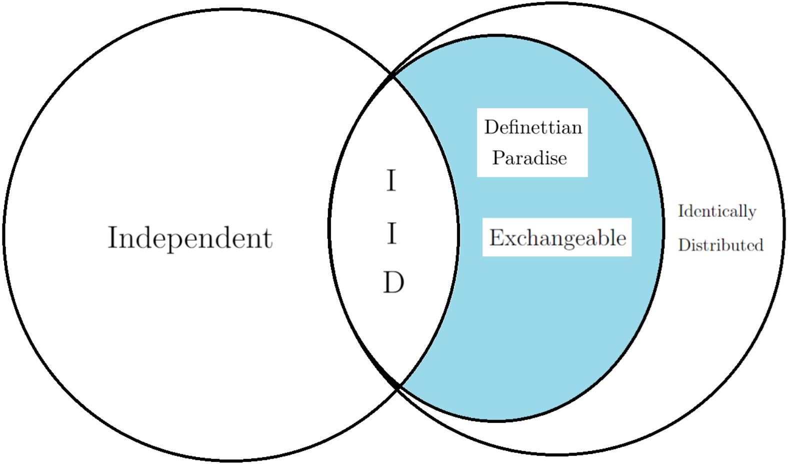
- If (U_1, U_2) is uniformly distributed on $\{(0,1), (1,2), (2,0)\}$, then U_1 and U_2 are Identically Distributed, but U_1 and U_2 are not Exchangeable, since $P(U_1 < U_2) = 2/3$, while $P(U_2 < U_1) = 1/3$.

Exchangeability to the end

Insper



Exchangeability to the end (Bayesian interlude)



Exchangeability to the end

- Let $W, U_1, U_2, \ldots$ be IID Bernoulli(1/2)

Exchangeability to the end

Insper

- Let $W, U_1, U_2, \ldots$ be IID Bernoulli(1/2)

- Defining $Z_i = U_i + W$, the Z_i's are not independent

Exchangeability to the end

- Let $W, U_1, U_2, \ldots$ be IID Bernoulli(1/2)
- Defining $Z_i = U_i + W$, the Z_i's are not independent
- $P\{Z_1 = 2\} = P\{U_1 = 1, W = 1\} = 1/4$

Exchangeability to the end

Insper

- Let $W, U_1, U_2, \dots$ be IID Bernoulli(1/2)
- Defining $Z_i = U_i + W$, the Z_i's are not independent
- $P\{Z_1 = 2\} = P\{U_1 = 1, W = 1\} = 1/4$
- $P\{Z_1 = 2 \mid Z_2 = 0\} = 0$

Exchangeability to the end

- Let $W, U_1, U_2, \ldots$ be IID Bernoulli(1/2)
- Defining $Z_i = U_i + W$, the Z_i's are not independent
- $P\{Z_1 = 2\} = P\{U_1 = 1, W = 1\} = 1/4$
- $P\{Z_1 = 2 \mid Z_2 = 0\} = 0$
- But $\{Z_i\}_{i \geq 1}$ is exchangeable

Exchangeability to the end

Insper

$$
\begin{aligned}
P(Z_1 = z_1, \ldots, Z_n = z_n)
&= \sum_{w \in \{0,1\}} P(Z_1 = z_1, \ldots, Z_n = z_n \mid W = w) P(W = w) \\
&= \sum_{w \in \{0,1\}} P(U_1 = z_1 - w, \ldots, U_n = z_n - w \mid W = w) P(W = w) \\
&= \sum_{w \in \{0,1\}} P(U_1 = z_1 - w, \ldots, U_n = z_n - w) P(W = w) \\
&= \sum_{w \in \{0,1\}} P(U_{\pi(1)} = z_1 - w, \ldots, U_{\pi(n)} = z_n - w) P(W = w) \\
&= \sum_{w \in \{0,1\}} P(U_{\pi(1)} = z_1 - w, \ldots, U_{\pi(n)} = z_n - w \mid W = w) P(W = w) \\
&= \sum_{w \in \{0,1\}} P(Z_{\pi(1)} = z_1, \ldots, Z_{\pi(n)} = z_n \mid W = w) P(W = w) \\
&= P(Z_{\pi(1)} = z_1, \ldots, Z_{\pi(n)} = z_n)
\end{aligned}
$$

Exchangeability to the end

- The most general characterization of exchangeable sequences is given by de Finetti's representation theorem.

Exchangeability to the end

Insper

- The most general characterization of exchangeable sequences is given by de Finetti's representation theorem.

- For a sequence $\{Z_i\}_{i\geq 1}$ of random variables taking values in $\{0,1\}$, de Finetti's representation theorem states that $\{Z_i\}_{i\geq 1}$ is exchangeable if and only if there is a random variable $\Theta : \Omega \to [0,1]$ such that, given that $\Theta = \theta$, the random variables $\{Z_i\}_{i\geq 1}$ are conditionally independent and identically distributed with distribution Bernoulli(θ).

Exchangeability to the end

Insper

- The most general characterization of exchangeable sequences is given by de Finetti's representation theorem.
- For a sequence $\{Z_i\}_{i\geq 1}$ of random variables taking values in $\{0,1\}$, de Finetti's representation theorem states that $\{Z_i\}_{i\geq 1}$ is exchangeable if and only if there is a random variable $\Theta : \Omega \to [0,1]$ such that, given that $\Theta = \theta$, the random variables $\{Z_i\}_{i\geq 1}$ are conditionally independent and identically distributed with distribution Bernoulli(θ).
- Furthermore, the distribution μ_Θ of Θ is unique, and $(1/m)\sum_{i=1}^m Z_i$ converges almost surely to Θ, when m tends to infinity.

Exchangeability to the end

Insper

Proposition 1. *Under the data exchangeability assumption, the sequence of conformity scores* $\{R_i\}_{i\geq 1}$ *is exchangeable.*

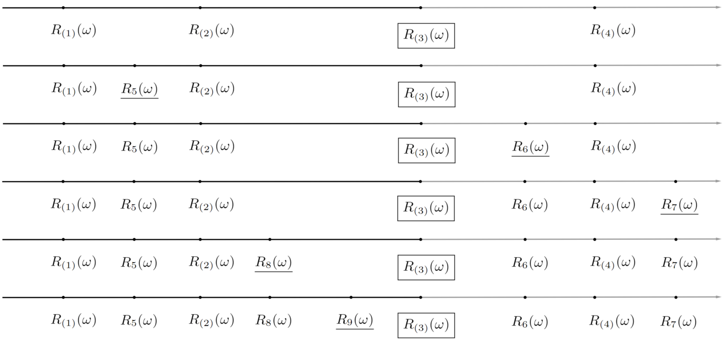
Exchangeability to the end

Insper

Proposition 1. *Under the data exchangeability assumption, the sequence of conformity scores* $\{R_i\}_{i\geq 1}$ *is exchangeable.*

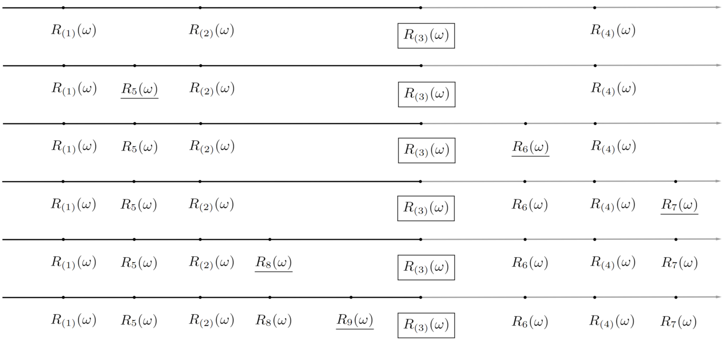
Proposition 3. *For exchangeable data, if the conformity scores are almost surely distinct, the sequence of indicators* $\{Z_i\}_{i\geq 1}$ *is exchangeable.*

Exchangeability to the end



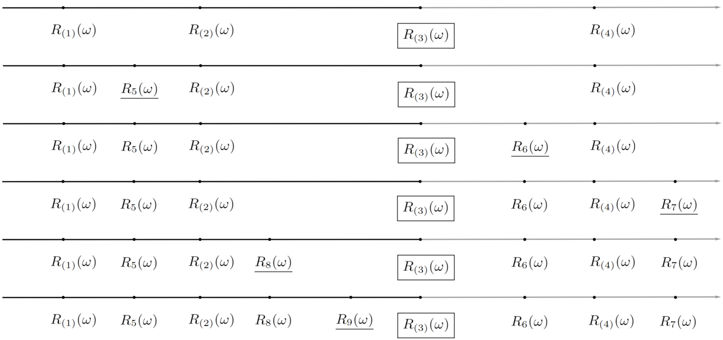
Consider the particular case depicted in the figure above, in which we have a calibration sample of size $n = 4$, the nominal miscoverage level $\alpha = 0.45$ (so that $\lceil (1-\alpha)(n+1) \rceil = 3$), and the horizon $m = 5$.

Exchangeability to the end



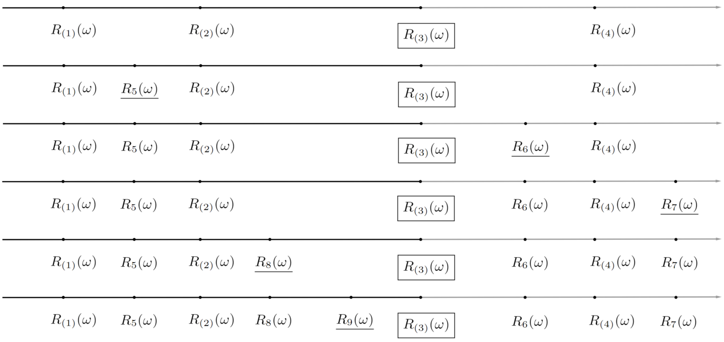
Let $b = \lceil (1-\alpha)(n+1) \rceil$ and $g = n - \lceil (1-\alpha)(n+1) \rceil + 1 = \lfloor \alpha(n+1) \rfloor$.

Exchangeability to the end



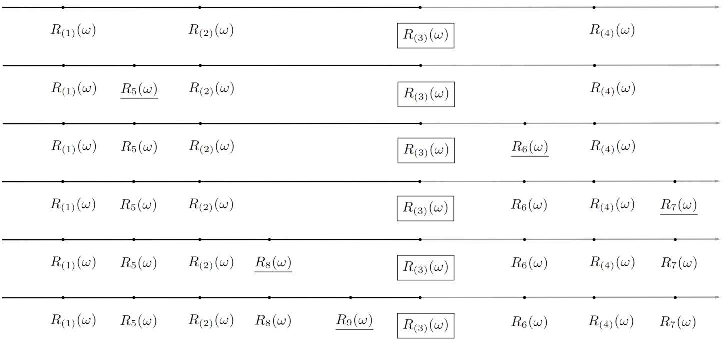
Since the sequence of conformity scores is exchangeable, R_5 has the same probability of falling into one of the $b + g = n + 1 = 5$ intervals defined by the ordered calibration conformity scores $R_{(1)}, \ldots, R_{(4)}$ (first line in the figure).

Exchangeability to the end



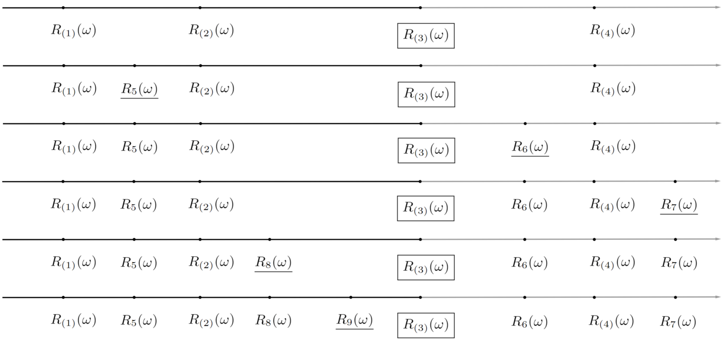
Remember that $Z_i = 1$ if an only if $R_i < R_{(\lceil (1-\alpha)(n+1) \rceil)}$.

Exchangeability to the end



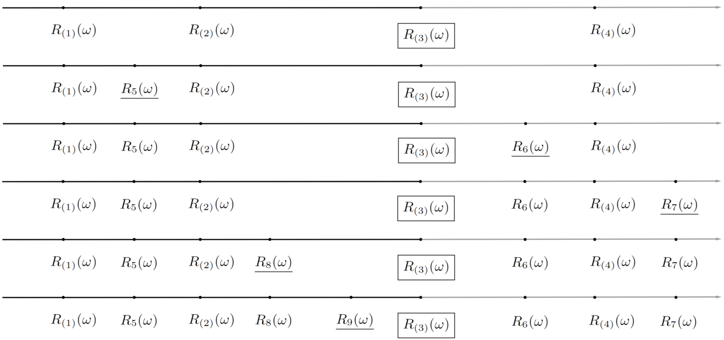
Hence, $Z_1 = 1$ if and only if R_5 falls into one of the $b = 3$ black intervals to the left of $R_{(3)}$ (second line in the figure), yielding $P(Z_1 = 1) = b/(n+1)$.

Exchangeability to the end



Given that $Z_1 = 1$, R_6 has, by exchangeability, the same probability of falling into one of the $(b+1)+g = n+2 = 6$ intervals defined the conformity scores $R_{(1)}, \ldots, R_{(4)}, R_5$, and $Z_2 = 0$ if and only if R_6 falls into one the $g = 2$ gray intervals to the right of $R_{(3)}$ (third line in the figure). Therefore, $P(Z_2 = 0 \mid Z_1 = 1) = g/(n+2)$.

Exchangeability to the end



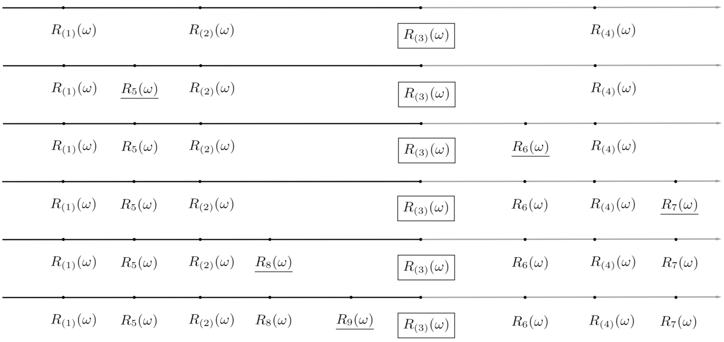
Following this reasoning, the product rule yields

$$P(Z_1 = 1, Z_2 = 0, Z_3 = 0, Z_4 = 1, Z_5 = 1) = \frac{b}{n+1} \cdot \frac{g}{n+2} \cdot \frac{g+1}{n+3} \cdot \frac{b+1}{n+4} \cdot \frac{b+2}{n+5},$$

which is manifestly exchangeable, as expected.

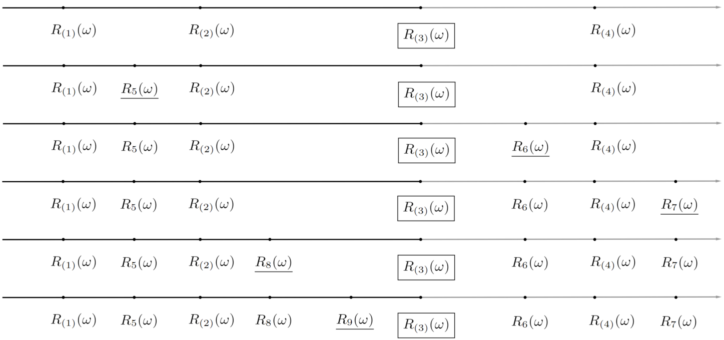
Exchangeability to the end

Insper



This is a Pólya's urn scheme with outcome `BGGBB` in which we started with b black balls (`B`) and g gray balls (`G`), and after drawing a ball from the urn we put it back adding one ball of the same color.

Exchangeability to the end



The exchangeability of the vector of indicators $(Z_1, Z_2, Z_3, Z_4, Z_5)$ implies that the event $\{Z_1 + Z_2 + Z_3 + Z_4 + Z_5 = 2\}$ is the union of $\binom{m}{2}$ mutually exclusive and equiprobable events of the form
$\{Z_1 = z_1, Z_2 = z_2, Z_3 = z_3, Z_4 = z_4, Z_5 = z_5\}$, in which 2 of the z_i's are equal to 1, and $m - 2 = 3$ of the z_i's are equal to 0. Therefore,

$$P(Z_1 + Z_2 + Z_3 + Z_4 + Z_5 = 2) = \binom{m}{2}\left(\frac{b}{n+1}\cdot\frac{g}{n+2}\cdot\frac{g+1}{n+3}\cdot\frac{b+1}{n+4}\cdot\frac{b+2}{n+5}\right).$$

Exchangeability to the end

Insper

Theorem 1. *Under the data exchangeability assumption, for every nominal miscoverage level $0 < \alpha < 1$, every calibration sample size $n \geq 1$, and every horizon $m \geq 1$, if the conformity scores are almost surely distinct, the distribution of the future coverage is given by*

$$P(C_m^{(n,\alpha)} = k/m) = \binom{m}{k} \frac{\left(\prod_{i=1}^{k}(\lceil(1-\alpha)(n+1)\rceil + i - 1)\right)\left(\prod_{i=1}^{m-k}(\lfloor\alpha(n+1)\rfloor + i - 1)\right)}{\prod_{i=1}^{m}(n+i)},$$

for $k = 1, \ldots, m$.

Exchangeability to the end

Insper

Theorem 1. *Under the data exchangeability assumption, for every nominal miscoverage level $0 < \alpha < 1$, every calibration sample size $n \geq 1$, and every horizon $m \geq 1$, if the conformity scores are almost surely distinct, the distribution of the future coverage is given by*

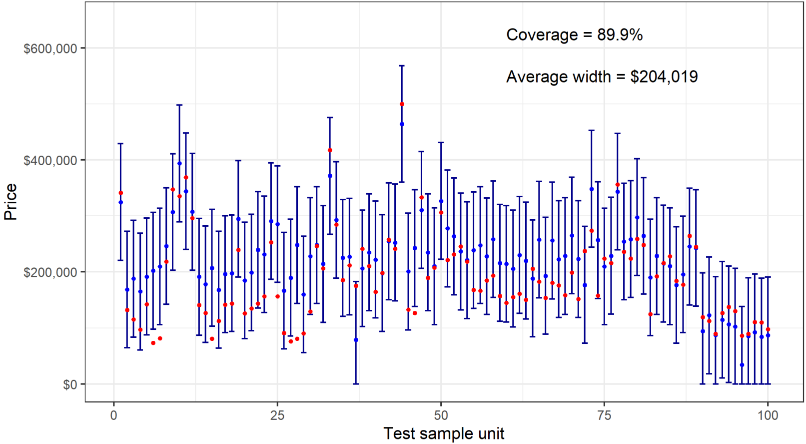
$$P(C_m^{(n,\alpha)} = k/m) = \binom{m}{k} \frac{\left(\prod_{i=1}^{k}(\lceil(1-\alpha)(n+1)\rceil + i - 1)\right)\left(\prod_{i=1}^{m-k}(\lfloor\alpha(n+1)\rfloor + i - 1)\right)}{\prod_{i=1}^{m}(n+i)},$$

for $k = 1, \ldots, m$.

Theorem 2. *For exchangeable data, if the conformity scores are almost surely distinct, the future coverage $C_m^{(n,\alpha)}$ converges almost surely when the horizon m tends to infinity to a random variable $C_\infty^{(n,\alpha)}$ with distribution $\mathrm{Beta}(\lceil(1-\alpha)(n+1)\rceil, \lfloor\alpha(n+1)\rfloor)$, for every nominal miscoverage level $0 < \alpha < 1$, and every calibration sample size $n \geq 1$.*

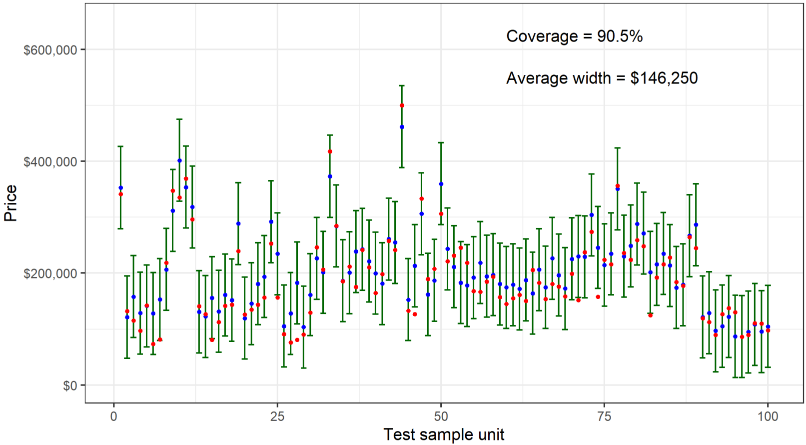
Linear regression standard score

Insper



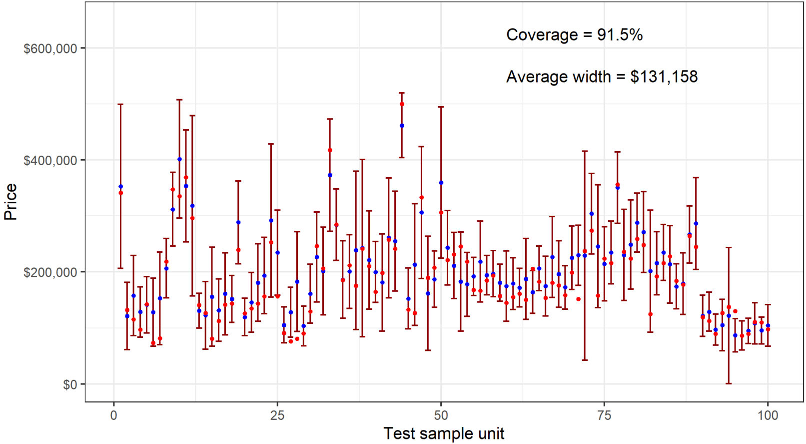
Random Forest standard score

Insper

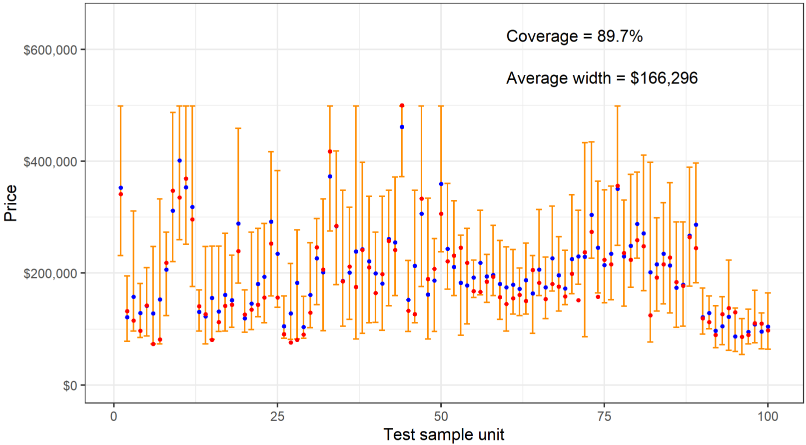


Random Forest locally-weighted score

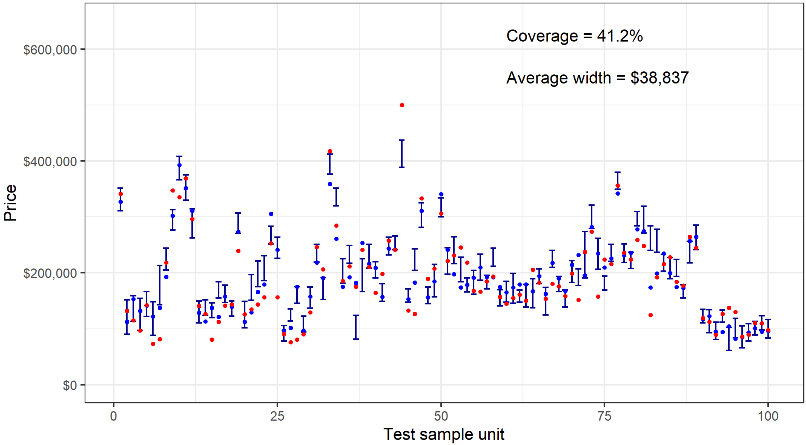
Insper



Conformalized Quantile Regression

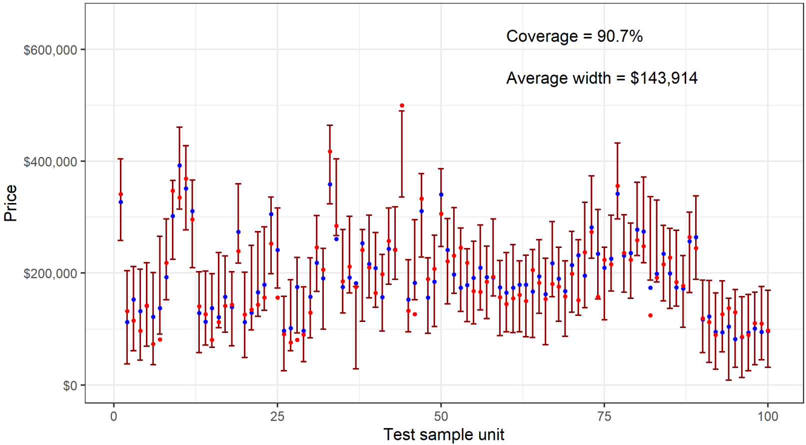


Bayesian Additive Regression Trees (BART)

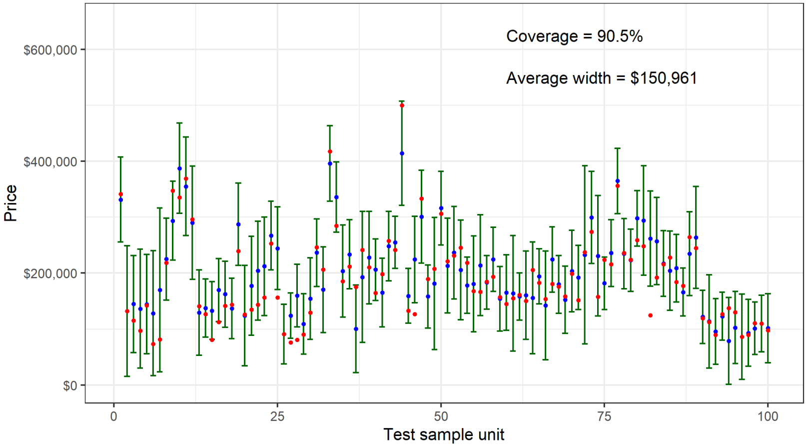


Conformalized BART (version 1)

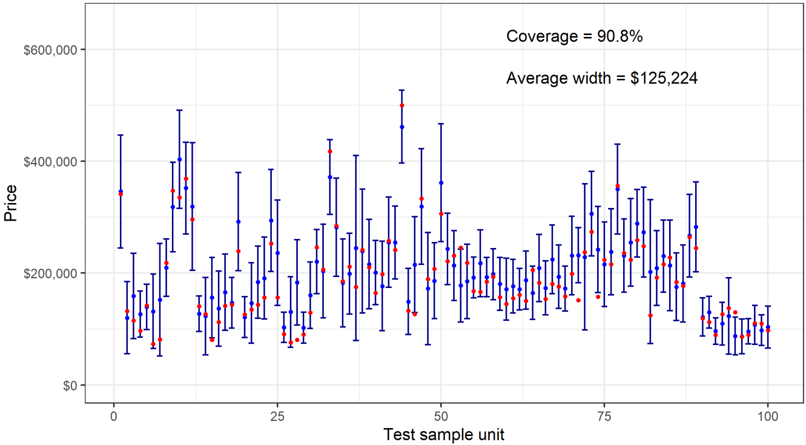
Insper



Conformalized BART (version 2)



Mysterious Conformal Prediction



Happy holidays!

Insper