Semiparametric Conformal Prediction   arXiv: 2411.02114

Ji Won Park[1] (park.ji_won@gene.com), Robert Tibshirani[2], and Kyunghyun Cho[1, 3]

[1] Frontier Research, Prescient Design, Genentech, [2] Department of Statistics, Stanford University, [3] Department of Computer Science, New York University

Summary

- Conformal prediction (CP) delivers valid prediction sets for any underlying predictor.
- Many applications, including drug design, require well-calibrated prediction sets over multiple correlated target variables.
- The CP algorithm involves extracting the empirical $1-\alpha$ quantile of *non-conformity scores*, where α is the miscoverage rate.
- We model the joint correlation structure of the scores using a **nonparametric vine copula** and improve the asymptotic efficiency of its plug-in quantile using **semiparametric one-step estimation**.

Problem: multi-target regression

Given a dataset $\{(X^{(i)}, Y^{(i)})\}_{i \in \mathcal{I}}$ of input features $X^{(i)} \in \mathcal{X}$ and vector-valued labels $Y^{(i)} \in \mathbb{R}^d$ and a new test point X^*, determine the set $\Gamma_{1-\alpha}(X^*) \subset \mathbb{R}^d$ that is **marginally valid** at level $1-\alpha$:

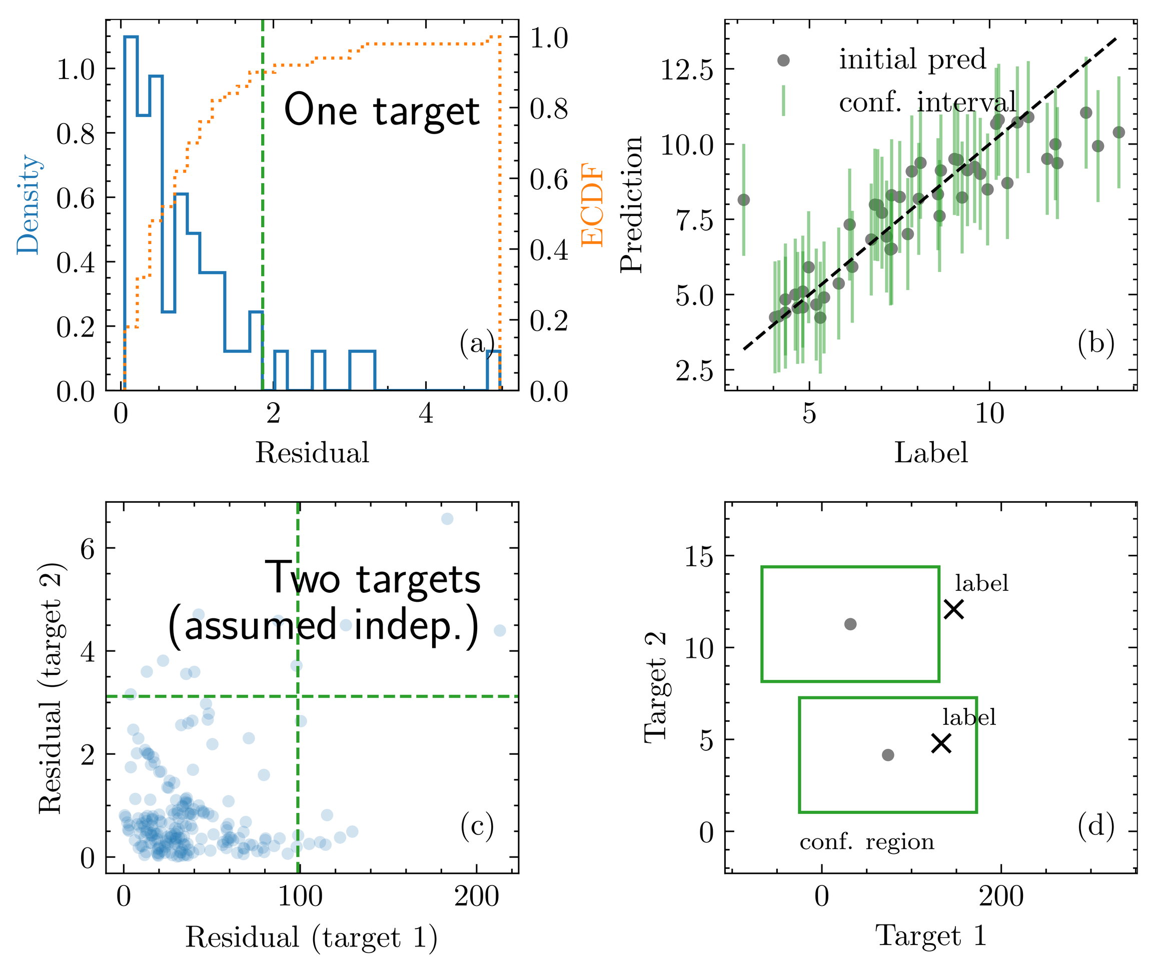
$$\mathbb{P}[Y^* \in \Gamma_{1-\alpha}(X^*)] \geq 1-\alpha.$$

Conformal prediction

Standard split CP.

1. Split the labeled dataset into the training data used to fit the underlying predictor $\hat{f}$ and the calibration data used to conformalize it.
2. Define the *non-conformity score* $S(X, Y, \hat{f})$ measuring how much (X, Y) deviates from the patterns learned by $\hat{f}$.
3. Evaluate the calibration scores and extract their $1-\alpha$ quantile, $Q_{1-\alpha}$.
4. At test time, gather outcomes yielding the smallest scores:

$$\Gamma_{1-\alpha}(X^*) = \{Y \in \mathbb{R} : S(X^*, Y, \hat{f}) \preccurlyeq Q_{1-\alpha}\}.$$



Semiparametric Conformal Prediction

Nonparametric vine copulas

Decompose d-variate distribution of score vector S into:

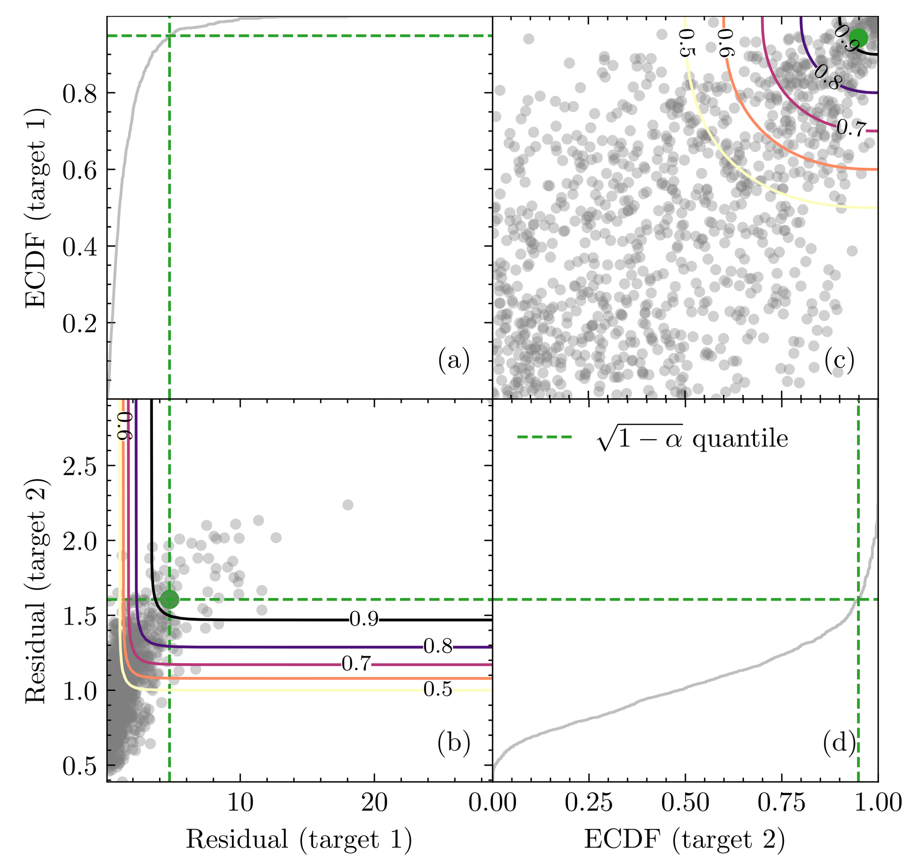
- **marginal CDF** of each component S_j, $j \in [d]$

$$F_j(s_j) = \mathbb{P}[S_j \leq s_j]$$

- **copula** $C: [0,1]^d \to [0,1]$, the joint CDF of uniform margins

$$C(u_1, \ldots, u_d) = \mathbb{P}[F_1(S_1) \leq u_1, \ldots, F_d(S_d) \leq u_d]$$

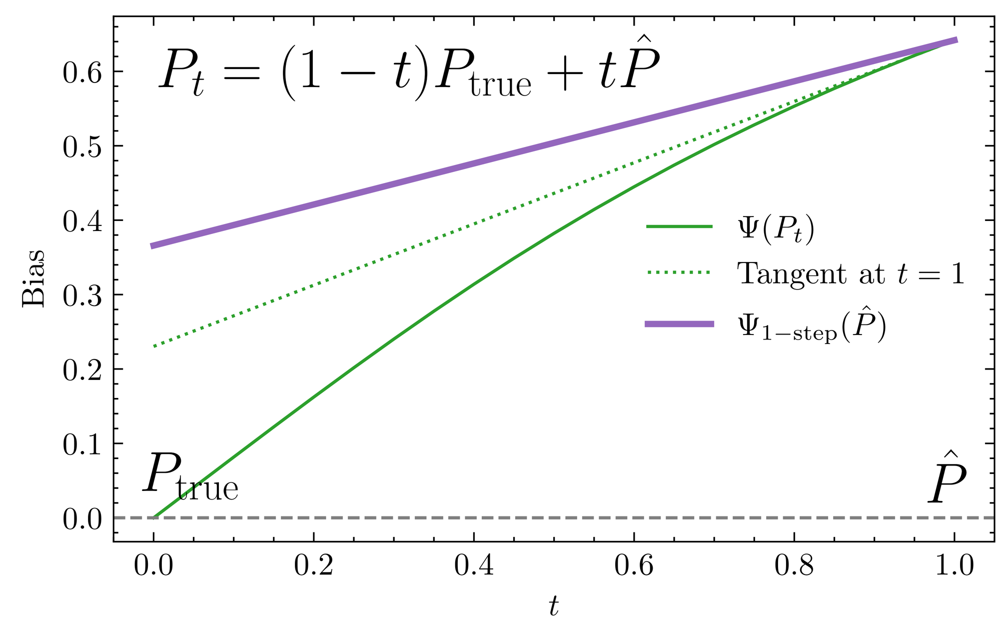
The **vine copula** (sequence of bivariate copula blocks) can accommodate arbitrarily complex dependencies.



Semiparametric one-step correction

The **efficient influence function** $\phi_P(S^{(i)})$ says how sensitive the plug-in parameter $\Psi(P)$ is to changes in the distribution P. The **one-step estimator** is asymptotically efficient:

$$\Psi_{\text{1-step}}(\hat{P}) = \Psi(\hat{P}) + \frac{1}{n}\sum_{i=1}^{n} \phi_{\hat{P}}(S^{(i)}).$$



Theoretical guarantees

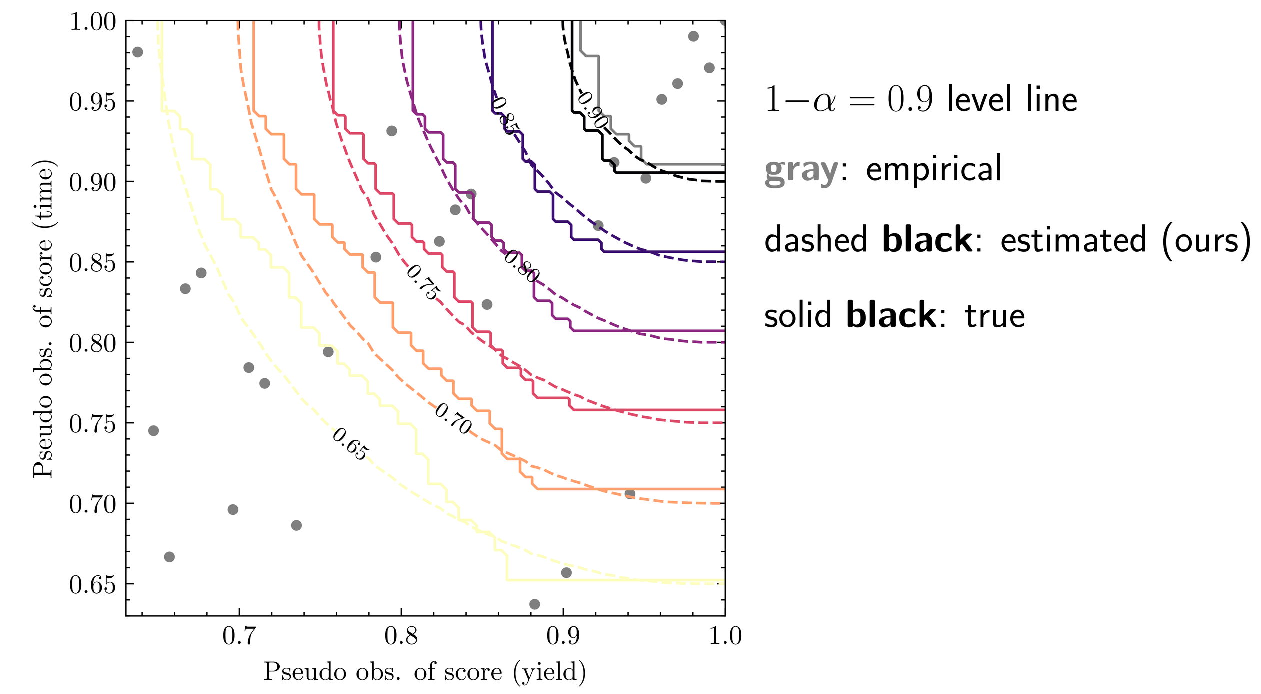
- **Asymp. exact coverage**: $\mathbb{P}[Y^* \in \Gamma_{1-\alpha}(X^*)] \to 1-\alpha$ as $n \to \infty$
- **Finite-sample approx. validity**: $\mathbb{P}[Y^* \in \Gamma_{1-\alpha}(X^*)] \geq 1-\alpha-\epsilon$, where ϵ is the TV distance between true and estimated distributions

Experimental results

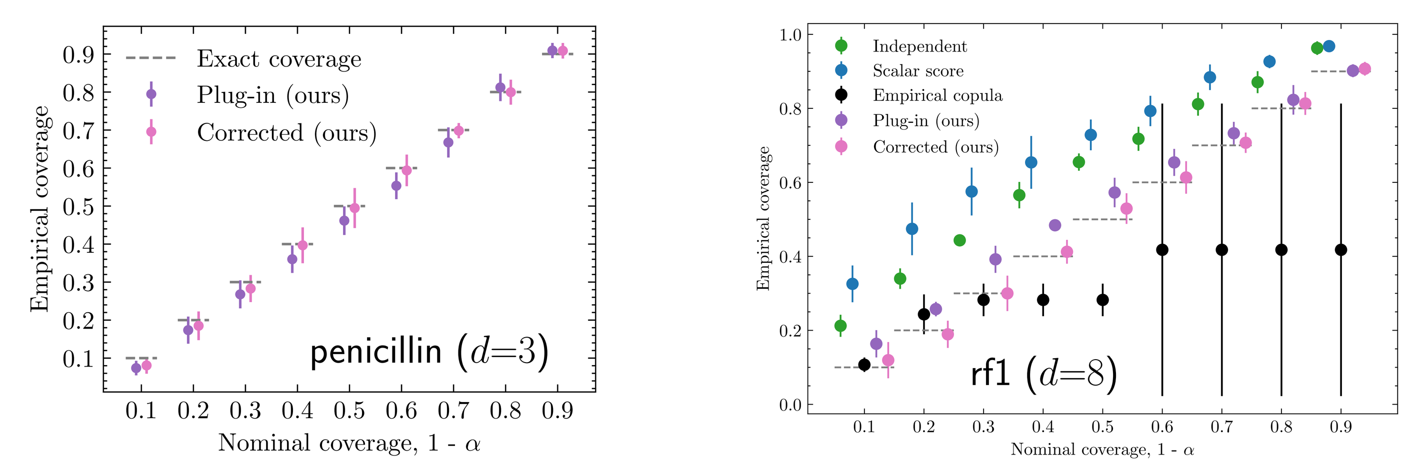
Joint modeling improves precision.

Method	Stock ($d=6, n=63$) Coverage	Stock Efficiency ↓	Caco2+ ($d=6, n=137$) Coverage	Caco2+ Efficiency ↓	rf1 ($d=8, n=225$) Coverage	rf1 Efficiency ↓
Independent	0.90 ± 0.01	-2.4 ± 0.2	0.95 ± 0.01	12.2 ± 0.1	0.96 ± 0.01	29.2 ± 0.5
Scalar score	0.90 ± 0.01	-1.8 ± 0.3	0.92 ± 0.01	28.3 ± 0.1	0.92 ± 0.00	27.7 ± 0.3
Empirical copula	0.50 ± 0.05	-4.7 ± 0.5	0.42 ± 0.08	8.4 ± 0.4	0.42 ± 0.12	21.2 ± 2.6
Plug-in (ours)	0.87 ± 0.02	-2.9 ± 0.2	0.90 ± 0.01	11.0 ± 0.1	0.91 ± 0.01	25.0 ± 0.2
Corrected (ours)	0.90 ± 0.02	-2.8 ± 0.2	0.93 ± 0.01	11.5 ± 0.2	0.91 ± 0.01	25.1 ± 0.3

Method	rf2 ($d=8, n=225$) Coverage	rf2 Efficiency ↓	scm1d ($d=16, n=448$) Coverage	scm1d Efficiency ↓	scm20d ($d=16, n=448$) Coverage	scm20d Efficiency ↓
Independent	0.96 ± 0.01	29.2 ± 0.5	0.96 ± 0.01	114.1 ± 0.4	0.96 ± 0.01	114.1 ± 0.4
Scalar score	0.92 ± 0.01	27.7 ± 0.3	0.89 ± 0.01	109.0 ± 0.3	0.89 ± 0.01	109.0 ± 0.3
Empirical copula	0.42 ± 0.12	21.2 ± 2.6	0.75 ± 0.05	108.5 ± 0.8	0.75 ± 0.05	108.5 ± 0.8
Plug-in (ours)	0.90 ± 0.01	25.0 ± 0.2	0.92 ± 0.01	111.4 ± 0.1	0.92 ± 0.01	111.4 ± 0.2
Corrected (ours)	0.91 ± 0.01	25.1 ± 0.3	0.91 ± 0.01	110.7 ± 0.2	0.90 ± 0.01	110.6 ± 0.2



One-step correction improves coverage at all α.



Our method handles missingness at random.

Under MAR, only taking the fully observed subset leads to biased quantiles. With an appropriate choice of copula model, we can natively impute missing observations.

Main references

- Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *JMLR* (2008).
- Gery Geenens. Probit transformation for kernel density estimation on the unit interval. *JASA* (2014).
- Lucien Le Cam. On the asymptotic theory of estimation and testing hypotheses. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (1956).