Dynamic Conformal Prediction for Multi-Target Regression

Optimising Informational Efficiency under Joint Validity

Filip Schlembach Evgueni Smirnov Mark H. M. Winands

Department of Advanced Computing Sciences
Maastricht University

September 12, 2025

Outline

Problem Setting

Inductive Conformal Prediction (ICP)

- Split a given sample $\mathcal{D} = \wr(\mathbf{x}^{(n)}, \mathbf{y}^{(n)})\int_{n=1}^{N}$ of N examples into a proper training set $\mathcal{D}_{\mathrm{tr}} = \wr(\mathbf{x}^{(n)}, \mathbf{y}^{(n)})\int_{n=1}^{N_{\mathrm{tr}}}$ of size N_{tr}, and a calibration set $\mathcal{D}_{\mathrm{cal}} = \wr(\mathbf{x}^{(n)}, \mathbf{y}^{(n)})\int_{n=N_{\mathrm{tr}}+1}^{N}$ of size $N_{\mathrm{cal}} = N - N_{\mathrm{tr}}$.
- Fit a model on $\mathcal{D}_{\mathrm{tr}}$ and obtain a point predictor $\hat{\mathbf{f}} : \mathbf{X} \to \mathbf{Y},\ \mathbf{x} \mapsto \hat{\mathbf{f}}(\mathbf{x}) = \hat{\mathbf{y}}$.
- Compute the nonconformity scores $\alpha^{(n)} = A\big((\mathbf{x}^{(n)}, \mathbf{y}^{(n)}), \hat{\mathbf{f}}\big)$ for each joint observation in the calibration set $\mathcal{D}_{\mathrm{cal}}$ using the nonconformity measure A.
- For a new object $\mathbf{x}^{(N+1)} \in \mathbf{X}$, return the prediction region

$$\Gamma^{\epsilon}(\mathbf{x}^{(N+1)}) = \left\{\tilde{\mathbf{y}} \in \mathbf{Y} : A\big((\mathbf{x}^{(N+1)}, \tilde{\mathbf{y}}), \hat{\mathbf{f}}\big) \leq \alpha_{(1-\epsilon)}\right\}. \tag{1}$$

Inductive conformal predictors are valid, i.e. $\mathbb{P}\big(\mathbf{y}^{(N+1)} \in \Gamma^{\epsilon}(\mathbf{x}^{(N+1)})\big) \geq 1 - \epsilon$ [8].

ICP for Univariate Regression

If $y \in Y \subseteq \mathbb{R}$

$$\Gamma^{\epsilon}(\mathbf{x}^{(N+1)}) = \left\{ y \in Y : A((\mathbf{x}^{(N+1)}, y), \hat{f}\ ) \leq \alpha_{(1-\epsilon)} \right\}. \tag{2}$$

is usually easy to solve and has a unique solution.

Example

Using absolute residuals $\alpha^{(n)} = |\hat{y}^{(n)} - y^{(n)}|$ as the NCM produces the prediction region

$$\Gamma^{\epsilon}(\mathbf{x}) = \left[\hat{y} - \alpha_{(1-\epsilon)},\ \hat{y} + \alpha_{(1-\epsilon)} \right], \tag{3}$$

where $\alpha_{(1-\epsilon)} = \mathrm{Q}_{1-\epsilon}(\{\alpha^{(n)}\}_{n=N_{\mathrm{tr}}+1}^{N})$ is the empirical $(1-\epsilon)$ quantile of $\{\alpha^{(n)}\}_{n=N_{\mathrm{tr}}+1}^{N}$.

ICP for Multivariate Regression

If $\mathbf{y} \in \mathbf{Y} \subseteq \mathbb{R}^M$

$$\Gamma^{\epsilon}(\mathbf{x}^{(N+1)}) = \Big\{ \tilde{\mathbf{y}} \in \mathbf{Y} : A\big((\mathbf{x}^{(N+1)}, \tilde{\mathbf{y}}), \hat{\mathbf{f}}\big) \leq \alpha_{(1-\epsilon)} \Big\}. \tag{4}$$

we observe two strategies in the literature.

Single-test methods employ an NCM $A : \mathbf{X} \times \mathbf{Y} \rightarrow \mathbb{R}$ [2, 3, 5].
The shape of the prediction region is determined by the NCM.

Multi-tests methods combine M single-target conformal predictors $\{\Gamma_m^{\epsilon_m}\}_{m=1}^M$, one for each dimension m of the label space, into a single Γ^{ϵ}.
These $\{\Gamma_m^{\epsilon_m}\}_{m=1}^M$ employ an NCM $A : \mathbf{X} \times \mathbb{R} \rightarrow \mathbb{R}$ which provides nonconformity scores for the associated dimension.

Prediction Regions of Multi-Tests ICP Methods

The prediction region of multi-tests methods is the Cartesian product of the M one-dimensional intervals constructed by the $\Gamma_m^{\epsilon_m}$ and therefore of hyperrectangular shape.

Benefits:

- Easy to interpret
- Prediction interval for one dimension is independent of the other dimensions

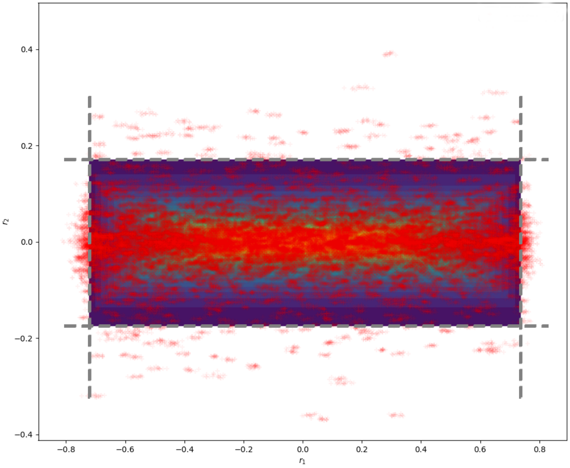


Figure: Example of 2D prediction region

Validity of Multi-Tests ICP Methods

We define validity for multi-tests methods as

$$\mathbb{P}\big(\mathbf{y}^{(N+1)} \in \Gamma^{\epsilon_h}(\mathbf{x}^{(N+1)})\big) = \mathbb{P}\left(\bigcap_{m=1}^{M} \left(y_m^{(n+1)} \in \Gamma_m^{\epsilon_m}(\mathbf{x}^{(N+1)})\right)\right) \geq 1 - \epsilon_h \; , \tag{5}$$

where $\Gamma^{\epsilon}(\mathbf{x}^{(N+1)})$ is the hyperrectangular prediction region with a global targeted error rate ϵ_h and $\Gamma_m^{\epsilon_m}(\mathbf{x}^{(N+1)})$ is the prediction interval for dimension m with a local targeted error rate ϵ_m [6].

Problem

How to select the local significance levels ϵ_m such that the resulting prediction region is valid and efficient?

Dynamic CP for Multi-Target Regression (DCP-MT)

Validity of Multi-Tests ICP Methods

Theorem

Let ϵ_h be the global targeted error rate for a multi-tests conformal predictor and let $\{\epsilon_m\}_{m=1}^{M}$ be the local error rates associated with each dimension of the label space. A multi-tests conformal predictor is valid, if

$$\epsilon_h \geq \sum_{m=1}^{M} \epsilon_m \ . \tag{6}$$

The result follows directly from Boole's inequality.

DCP-MT | Idea

- Equation (6) shows that for Γ^{ϵ_h} to be valid, it is sufficient that $\epsilon_h \geq \sum_{m=1}^{M} \epsilon_m$.
- We interpret ϵ_h as a global error budget.
- DCP-MT solves the problem of optimally allocating the global error budget ϵ_h between the M dimensions of the label space.

DCP-MT | Formally

Minimise the volume of the resulting hyperrectangular prediction region

$$\prod_{m=1}^{M} 2\mathrm{Q}_{1-\epsilon_m}\left(\{\alpha^{(m,n)}\}_{n=N_{\mathrm{tr}}+1}^{N}\right) \tag{7}$$

under the constraint

$$\epsilon_h \geq \sum_{m=1}^{M} \epsilon_m \ , \tag{8}$$

where $\{\alpha^{(m,n)}\}_{n=N_{\mathrm{tr}}+1}^{N}$ are all nonconformity scores associated with dimension m.

DCP-MT | MILP Problem Formulation

Finding the optimal allocation of the error budget is equivalent to minimising the cost function

$$\sum_{m=1}^{M} \sum_{n=1}^{N_{\mathrm{cal}}} I_m^{(n)} \ln(\alpha_m^{(n)}) \tag{9}$$

under the constraints

$$\epsilon_h \geq \sum_{m=1}^{M} \epsilon_m \ , \tag{10}$$

and

$$\sum_{n=1}^{N_{\mathrm{cal}}} I_m^{(n)} = 1 \ , \ \forall m \in \{1, \ldots, M\} \ . \tag{11}$$

Related Work

Related Work

- **Bonferroni** predictors assign the same value $\epsilon_m = \epsilon_h / M$ to all local significance levels [7].
- **Copula CP** tries to capture the dependencies across the nonconformity scores' dimensions by fitting copulas to them [4].
- **Max-Rank** assigns each example in the calibration set its maximum rank over all dimensions. [6].

All these methods assign the same ϵ_m to all dimensions.

Experimental Results & Discussion

2D Example | Uniform and Chi-Squared Noise I

Data set $\mathbf{y} = B\mathbf{x} + \mathbf{e}$
with $\mathbf{y} \in \mathbb{R}^2$, and
the noise term
$e \sim [\mathcal{U}(-1,1), \chi^2_2]^\top$

Model linear regression

NCM absolute residuals

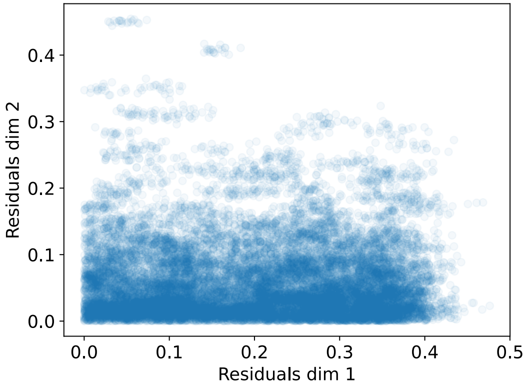


Figure: Absolute residuals

2D Example | Uniform and Chi-Squared Noise II

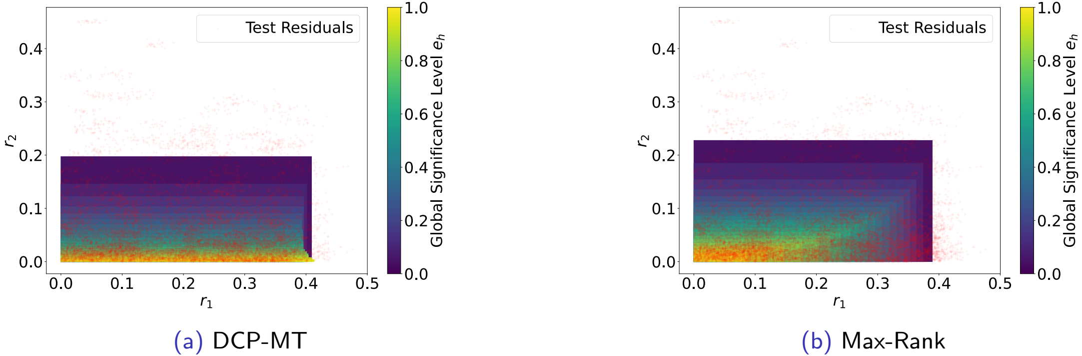
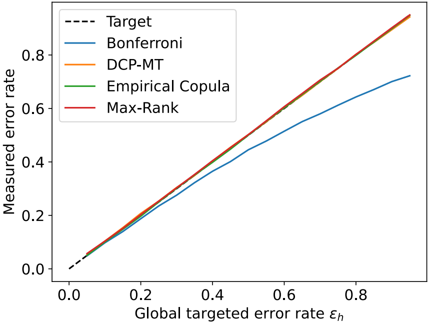
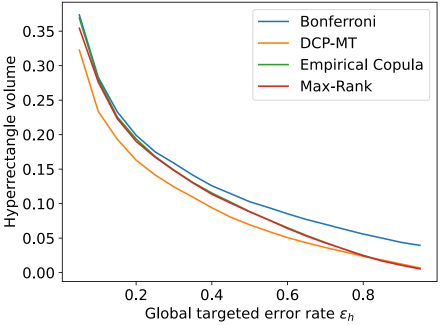


Figure: Multi-tests method's partition of the residual space for the synthetic data set with $e \sim [\mathcal{U}(-1,1), \chi^2_2]^\top$ averaged over 20 repetitions.



(a) Error rates



(b) Volumes

Figure: Multi-tests method's mean error rates and volumes for the synthetic data set with $e \sim [\mathcal{U}(-1,1), \chi^2_2]^\top$ over 20 repetitions.

Experimental Results I

Data Set	Bonferroni	DMT-CP	Empirical Copula	Max-Rank
synt, $e \sim [\mathcal{U}(-1, 1), \chi^2_2]^\top$	0.049 ± 0.006	**0.052 ± 0.005**	0.051 ± 0.006	0.057 ± 0.007
synt, $e \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma})$, $\sigma = 0$	0.045 ± 0.006	**0.054 ± 0.008**	0.045 ± 0.006	0.051 ± 0.008
synt, $e \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma})$, $\sigma = 0.5$	0.043 ± 0.004	**0.052 ± 0.005**	0.048 ± 0.004	0.051 ± 0.005
synt, $e \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma})$, $\sigma = 1$	0.023 ± 0.005	0.035 ± 0.005	**0.049 ± 0.006**	**0.049 ± 0.006**
synt, $e \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma})$, $\sigma = -0.5$	0.043 ± 0.004	**0.054 ± 0.006**	0.045 ± 0.005	0.051 ± 0.006
diabetes	0.035 ± 0.011	0.049 ± 0.013	0.038 ± 0.012	**0.050 ± 0.013**
music origin	0.042 ± 0.007	**0.047 ± 0.008**	0.048 ± 0.006	0.051 ± 0.007
rf1	0.037 ± 0.002	0.037 ± 0.002	0.049 ± 0.003	**0.051 ± 0.002**
rf2	0.037 ± 0.002	**0.036 ± 0.002**	0.048 ± 0.002	0.050 ± 0.002
scm1d	0.021 ± 0.001	0.026 ± 0.001	0.050 ± 0.002	**0.052 ± 0.002**
scm20d	0.021 ± 0.002	0.027 ± 0.002	0.049 ± 0.002	**0.051 ± 0.002**

Table: Error rates acoss all experiments for $\epsilon_h = 0.05$. Bold values indicate that the method produced the smallest prediction region while underlined values highlight instances for which the measured error rate ≥ 0.05.

Experimental Results II

Data Set	Bonferroni	DMT-CP	Empirical Copula	Max-Rank
synt, $e \sim [\mathcal{U}(-1,1), \chi_2^2]^\top$	0.374 ± 0.013	$\mathbf{0.323 \pm 0.011}$	0.369 ± 0.013	0.354 ± 0.011
synt, $e \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma})$, $\sigma = 0$	6.644 ± 0.246	$\mathbf{6.312 \pm 0.205}$	6.628 ± 0.232	6.347 ± 0.201
synt, $e \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma})$, $\sigma = 0.5$	4.390 ± 0.126	$\mathbf{4.124 \pm 0.114}$	4.263 ± 0.122	4.159 ± 0.105
synt, $e \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma})$, $\sigma = 1$	5.928 ± 0.305	5.682 ± 0.281	$\mathbf{4.552 \pm 0.221}$	$\mathbf{4.552 \pm 0.221}$
synt, $e \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma})$, $\sigma = -0.5$	6.068 ± 0.230	$\mathbf{5.621 \pm 0.150}$	5.962 ± 0.239	5.720 ± 0.230
diabetes	7.919 ± 0.943	6.604 ± 0.752	7.627 ± 0.923	$\mathbf{6.498 \pm 0.784}$
music origin	24.417 ± 2.147	$\mathbf{22.860 \pm 1.602}$	23.433 ± 1.969	23.237 ± 2.002
rf1	172.734 ± 81.099	32.901 ± 14.016	38.125 ± 18.702	$\mathbf{32.341 \pm 15.305}$
rf2	87.792 ± 46.649	$\mathbf{14.946 \pm 8.429}$	19.484 ± 12.388	16.539 ± 10.866
scm1d	11.387E9 $\pm$ 5.455E9	3.753E9 $\pm$ 1.420E9	41.820E6 $\pm$ 11.540E6	**32.774E6 $\pm$ 8.828E6**
scm20d	108.772E9 $\pm$ 36.307E9	35.955E9 $\pm$ 11.015E9	1.172E9 $\pm$ 0.345E9	**0.942E9 $\pm$ 0.301E9**

Table: Hyperrectangle volumes acoss all experiments for $\epsilon_h = 0.05$. Bold values indicate that the method produced the smallest prediction region while underlined values highlight instances for which the measured error rate ≥ 0.05.

Discussion

- DCP-MT produces smaller prediction regions than copula conformal predictors and Max-Rank for data sets whose nonconformity scores are weakly correlated across their dimensions and/or have differing marginal distributions.
- In the case of strong positive correlation between the dimensions' nonconformity scores Max-Rank excels. Here, DCP-MT is limited by Boole's inequality.
- DCP-MT has high computational cost and can be unstable.

Conclusion

Conclusion

- DCP-MT is a novel approach to determine the local targeted error rates of multiple single-target ICPs.
- DCP-MT minimises the hyperrectangle's volume by dynamically allocating a global error budget between the local targeted error rates using integer linear programming.
- DCP-MT is valid.
- DCP-MT produces smaller prediction regions than copula conformal predictors and Max-Rank for data sets whose nonconformity scores are weakly correlated across their dimensions and/or have differing marginal distributions.
- Other methods that are not bound by Boole's inequality work better when the nonconformity scores are strongly correlated across their dimensions.

Future Work

- Extend the approach of optimising the prediction region's volume to copula conformal predictors.
- Optimise the implementation to utilise monotonicity of the volume as a function of the targeted error rate in each dimension.
- Lift / soften the constrained imposed by Boole's inequality.
- Compare DCP-MT to the method proposed by Matthew Cleaveland et al. "Conformal Prediction Regions for Time Series Using Linear Complementarity Programming" [1].

Thank you

Bibliography I

[1] Matthew Cleaveland et al. "Conformal Prediction Regions for Time Series Using Linear Complementarity Programming". en. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 38.19 (Mar. 2024), pp. 20984–20992. ISSN: 2374-3468, 2159-5399. DOI: 10.1609/aaai.v38i19.30089. URL: https://ojs.aaai.org/index.php/AAAI/article/view/30089 (visited on 07/02/2025).

[2] Victor Dheur et al. *A Unified Comparative Study with Generalized Conformity Scores for Multi-Output Conformal Regression*. _eprint: 2501.10533. Feb. 2025. URL: https://arxiv.org/abs/2501.10533.

[3] Chancellor Johnstone and Bruce Cox. "Conformal uncertainty sets for robust optimization". In: *Proceedings of the Tenth Symposium on Conformal and Probabilistic Prediction and Applications*. Ed. by Lars Carlsson et al. Vol. 152. Proceedings of Machine Learning Research. PMLR, Sept. 2021, pp. 72–90. URL: https://proceedings.mlr.press/v152/johnstone21a.html.

[4] Soundouss Messoudi, Sébastien Destercke, and Sylvain Rousseau. "Copula-based conformal prediction for multi-target regression". en. In: *Pattern Recognition* 120 (June 2021), p. 108101. ISSN: 00313203. DOI: 10.1016/j.patcog.2021.108101. URL: https://linkinghub.elsevier.com/retrieve/pii/S0031320321002880 (visited on 02/10/2022).

[5] Soundouss Messoudi, Sébastien Destercke, and Sylvain Rousseau. "Ellipsoidal conformal inference for Multi-Target Regression". In: *Proceedings of the Eleventh Symposium on Conformal and Probabilistic Prediction with Applications*. Ed. by Ulf Johansson et al. Vol. 179. Proceedings of Machine Learning Research. PMLR, Aug. 2022, pp. 294–306. URL: https://proceedings.mlr.press/v179/messoudi22a.html.

[6] Alexander Timans et al. *Max-Rank: Efficient Multiple Testing for Conformal Prediction*. _eprint: 2311.10900. Mar. 2025. URL: https://arxiv.org/abs/2311.10900.

Bibliography III

[7] Vladimir Vovk. "Transductive conformal predictors". In: *Artificial Intelligence Applications and Innovations*. Ed. by Harris Papadopoulos et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 348–360. ISBN: 978-3-642-41142-7. DOI: 10.1007/978-3-642-41142-7_36.

[8] Vladimir Vovk, A. Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*. en. New York: Springer, 2005. ISBN: 978-0-387-00152-4 978-0-387-25061-8.