




# nonconform: Conformal Anomaly Detection (Python)

Oliver Hennhöfer <sup>1</sup>

<sup>1</sup> Intelligent Systems Research Group, Karlsruhe University of Applied Sciences, Karlsruhe, Germany

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

## Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Open Journals](#) 

Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

## Summary

The ability to quantify uncertainty is a fundamental requirement for AI systems in safety-critical and costly-to-error domains, as reliable decision-making strongly depends on it. The Python package nonconform offers statistically principled uncertainty quantification for semi-supervised anomaly detection based on one-class classification (Petsche & Gluck, 1994). The package implements methods from conformal anomaly detection (Bates et al., 2023; Jin & Candès, 2023; Laxhammar & Falkman, 2010), grounded in the principles of conformal inference (Lei & Wasserman, 2013; Papadopoulos et al., 2002; Vovk et al., 2005).

In essence, nonconform transforms raw anomaly scores from an anomaly detection model into statistically valid  $p$ -values by calibrating the model on a hold-out set  $\mathcal{D}_{\text{calib}}$  of normal data. For a new observation  $X_{n+1}$  with score  $\hat{s}(X_{n+1})$ , this is achieved by comparing it to the empirical distribution of calibration scores  $\hat{s}(X_i)$  for  $i \in \mathcal{D}_{\text{calib}}$ . The conformal  $p$ -value  $\hat{u}(X_{n+1})$  is then defined as the normalized rank of  $\hat{s}(X_{n+1})$  among the calibration scores (Liang et al., 2024):

$$\hat{u}(X_{n+1}) = \frac{|\{i \in \mathcal{D}_{\text{calib}} : \hat{s}(X_i) \leq \hat{s}(X_{n+1})\}|}{n}.$$

By framing anomaly detection as a sequence of statistical hypothesis tests, these  $p$ -values enable systematic control of the False Discovery Rate (FDR) (Bates et al., 2023; Benjamini & Hochberg, 1995) at a pre-defined significance level. The library integrates seamlessly with the widely used pyod library (Chen et al., 2024; Zhao et al., 2019), facilitating the application of conformal techniques across a broad range of anomaly detection models.

## Statement of Need

A central challenge in anomaly detection lies in setting an appropriate detection threshold, as it directly determines the false positive rate. In high-stakes domains such as fraud detection, medical diagnostics, and industrial quality control, controlling false positives is critical: excessive false alarms can cause *alert fatigue* and ultimately render a system impractical. The nonconform package addresses this issue by replacing raw anomaly scores with  $p$ -values, thereby enabling formal FDR control. As a result, the conformal methods become effectively *threshold-free*, since decision thresholds are determined by the underlying statistical procedures.

$$FDR = \frac{\text{Efforts Wasted on False Alarms}}{\text{Total Efforts}}$$

(Benjamini et al., 2009)

Moreover, conformal methods are *non-parametric* and *model-agnostic*, and thus apply to any model that produces consistent anomaly scores on arbitrarily distributed data. A key requirement of these methods is the statistical assumption of exchangeability between calibration and test data, which ensures the validity of conformal  $p$ -values. Exchangeability requires only

that the joint distribution of data is invariant under permutations, making it considerably more general (and thus less restrictive) than the independent and identically distributed (*i.i.d.*) assumption commonly imposed in classical machine learning.

To operationalize this assumption in practice, the `nonconform` package provides several strategies for constructing calibration sets from training data, including approaches tailored to low-data regimes (Hennhofer & Preisach, 2024), that do not rely on a dedicated hold-out set. Based on obtained calibration sets, the package derives either standard conformal  $p$ -values or weighted conformal  $p$ -values (Jin & Candès, 2023), which are particularly useful under covariate shift when exchangeability is only approximate. By offering these tools, `nonconform` enables researchers and practitioners to build anomaly detectors whose outputs are statistically controlled to maintain the FDR at a chosen nominal level.

Overall, the reliance on exchangeability makes these methods well suited for cross-sectional data, but less appropriate for time-series applications where temporal ordering carries essential information.

## Acknowledgements

This work was conducted in part within the research projects *Biflex Industrie* (grant number 01MV23020A) funded by the German Federal Ministry of Economic Affairs and Climate Action (BMWK).

## References

- Bates, S., Candès, E., Lei, L., Romano, Y., & Sesia, M. (2023). Testing for outliers with conformal  $p$ -values. *The Annals of Statistics*, 51(1). <https://doi.org/10.1214/22-aos2244>
- Benjamini, Y., Heller, R., & Yekutieli, D. (2009). Selective inference in complex research. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1906), 425–4271. <https://doi.org/10.1098/rsta.2009.0127>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300. <https://doi.org/10.2307/2346101>
- Chen, S., Qian, Z., Siu, W., Hu, X., Li, J., Li, S., Qin, Y., Yang, T., Xiao, Z., Ye, W., Zhang, Y., Dong, Y., & Zhao, Y. (2024). PyOD 2: A python library for outlier detection with LLM-powered model selection. *arXiv Preprint arXiv:2412.12154*.
- Hennhofer, O., & Preisach, C. (2024). Leave-One-Out-, Bootstrap- and Cross-Conformal Anomaly Detectors. *2024 IEEE International Conference on Knowledge Graph (ICKG)*, 110–119. <https://doi.org/10.1109/ICKG63256.2024.00022>
- Jin, Y., & Candès, E. J. (2023). Model-free selective inference under covariate shift via weighted conformal  $p$ -values. <https://api.semanticscholar.org/CorpusID:259950903>
- Laxhammar, R., & Falkman, G. (2010). Conformal prediction for distribution-independent anomaly detection in streaming vessel data. *Proceedings of the First International Workshop on Novel Data Stream Pattern Mining Techniques*, 47–55. <https://doi.org/10.1145/1833280.1833287>
- Lei, J., & Wasserman, L. (2013). Distribution-free prediction bands for non-parametric regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(1), 71–96. <https://doi.org/10.1111/rssb.12021>
- Liang, Z., Sesia, M., & Sun, W. (2024). Integrative conformal  $p$ -values for out-of-distribution testing with labelled outliers. *Journal of the Royal Statistical Society Series B: Statistical*

- 79      *Methodology*, 86(3), 671–693. <https://doi.org/10.1093/jrsssb/qkad138>
- 80      Papadopoulos, H., Proedrou, K., Vovk, V., & Gammerman, A. (2002). Inductive confidence  
81      machines for regression. In *Machine learning: ECML 2002* (pp. 345–356). Springer Berlin  
82      Heidelberg. [https://doi.org/10.1007/3-540-36755-1\\_29](https://doi.org/10.1007/3-540-36755-1_29)
- 83      Petsche, T., & Gluck, M. (1994). Workshop on novelty detection and adaptive system  
84      monitoring. *Advances in Neural Information Processing Systems (NIPS)*.
- 85      Vovk, V., Gammerman, A., & Shafer, G. (2005). *Algorithmic learning in a random world*.  
86      Springer-Verlag. ISBN: 0387001522
- 87      Zhao, Y., Nasrullah, Z., & Li, Z. (2019). PyOD: A python toolbox for scalable outlier detection.  
88      *Journal of Machine Learning Research*, 20(96), 1–7. [http://jmlr.org/papers/v20/19-011.](http://jmlr.org/papers/v20/19-011.html)  
89      [html](http://jmlr.org/papers/v20/19-011.html)

DRAFT