The Journal of Open Source Software

binomial_cis: A Python Package for Optimal Binomial Confidence Intervals

Joseph A. Vincent [1]

1 Department of Aeronautics and Astronautics, Stanford University

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Summary

binomial_cis is a Python package for computing confidence intervals for the probability of success parameter, p, of a binomial distribution.

Statement of Need

Constructing confidence intervals for an unknown probability success given samples of successes and failures is one of the most fundamental problems in statistical inference. Research into this question dates back at least to the 1930s with the work of Clopper and Pearson (Clopper & Pearson, 1934). One of the key culminations of this research effort were procedures given by (Eudey, 1949) (Lehmann & Romano, 2022) to construct uniformly most accurate (UMA) and uniformly most accurate unbiased (UMAU) confidence intervals. These intervals have desirable optimality properties, and provide much better inference at small sample sizes than other methods. The need for binomial_cis arose out of these methods not having an open-source implementation.

Research Usage

binomial_cis has been used to compute confidence intervals for the success rate of robots in simulated and real-world tasks (Vincent et al., 2024).

Acknowledgements

Financial support was provided by Toyota Research Institute.

References

Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, *26*(4), 404–413.

Eudey, M. W. (1949). *On the treatment of discontinuous random variables* [PhD thesis]. University of California - Berkeley.

Lehmann, E. L., & Romano, J. P. (2022). *Testing statistical hypotheses* (Vol. 4). Springer.

Vincent, J. A., Nishimura, H., Itkina, M., Shah, P., Schwager, M., & Kollar, T. (2024). How generalizable is my behavior cloning policy? A statistical approach to trustworthy performance evaluation. *arXiv Preprint arXiv:2405.05439*.