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Published: unpublished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. The binomial distribution represents the likelihood of observing k successes in n trials where the probability of success for each trial is p. For example, p may be the probability of a coin flip landing on heads, and k the number of heads we observe after n flips. One often does not know the value of p and wishes to estimate this value. After observing k successes in n trials, a confidence interval is a set, constructed based on k, n, that covers the unknown parameter p with some user-specified probability. The binomial_cis package computes confidence intervals that lower and/or upper bound p with a user-specified probability.

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). A foundational result for constructing binomial confidence intervals of minimal width was given by (Eudey, 1949) and is also formalized in (Lehmann & Romano, 2022). We refer to these intervals as *optimal binomial confidence intervals* and they have the property of being uniformly most accurate (UMA) and uniformly most accurate unbiased (UMAU). Practically, these intervals can provide better inference of p at small sample sizes ($n \leq 50$). The binomial_cis package is the first open-source implementation of these optimal binomial confidence intervals. In addition, this software provides worst-case analysis of the tightness for the confidence intervals, a feature that is not present in other software for binomial confidence intervals. Practically, this feature assists the user in understanding how many samples an experiment should have in order to meet a desired level of accuracy in inferring the value of p.

Comparison to Existing Software

There are many existing software packages for computing binomial confidence intervals. binomial_cis differs from the existing software in two ways:

1. We provide open-source implementations for the optimal binomial confidence intervals given by (Eudey, 1949) and formalized in (Lehmann & Romano, 2022).
2. We provide functionality for worst-case analysis of the tightness for the confidence intervals, which helps guide users on selecting the sample size for their experiments.

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*.