in their conclusion but do not seem to have used them in their paper. Furthermore their code allows the user to provide different priors for the equal-tailed interval, whereby the symmetry assumption behind the equal-tailed interval might be violated. The implementation from Matsen IV et al. [13] considers the situation where both Beta distributions have the same priors, i.e. $\alpha = \theta$ and $\beta = \varphi$. They also note issues with hypergeometrics in R, which we also found, which was our motivation for using Python. Finally, our implementation is the only one of the three that can provide the symbolic functions, as showcased in the Use of package section. We are not aware of other software which allows for calculation of both the HPD and the equal-tailed interval.

DISCUSSION

Our implementation has been demostrated with examples from epidemiology but can be used in any science where 2×2 contingency table-like data arise. For example, within the field of ecology, the response ratio is an index used in meta-analysis of ecological experiments. It is implied in Hedges et al. [6] and Lajeunesse [11] that only confidence intervals are considered for these, thus our method will allow the calculation of credible intervals for this measure. Another metric mentioned in Lajeunesse [12] is the odds ratio, which is similar to the relative risk in certain situations. Hedges et al. [6] and Lajeunesse [12] consider the distribution of this ratio to which our method is applicable. Another biological ratio we could consider examining is mutation frequencies. Significance tests in this area rely on ratios of mutation rates [10]. Indeed, Matsen IV et al. [13]'s work seems to have arisen from work on hypermutations. We have named *bayesint* with the intention that it will be expanded to contain other intervals with Bayesian origins.

CONCLUSION

We have created a Python package *bayesint* which calculates two types of credible interval for the ratio of beta distributions. We used the package to calculate the credible intervals of the relative risk of subjects innoculated or not being attacked by cholera and highlighted the strengths of our package compared with other available options. We believe our implementation to be a useful tool for analysis of 2×2 contingency tables.

Implementation and architecture

The *bayesint* package is freely available at GitHub (*https://github.com/PublicHealthEngland/bayesint*) and through the PyPI index of Python packages (*pypi.org/project/bayesint*). This manuscript concerns *bayesint* version 1.0.3, the package's documentation will be updated if future versions have different requirements.

Quality control

bayesint is extensively tested in Python 2.7.14 and also functions in Python 3.5.5 and Python 3.6.6.

(2) AVAILABILITY

OPERATING SYSTEM

Any operating system where Python 2.7.14 or better or 3.5.5 or better with the dependencies enumerated below are installed. In practice this includes all three major operating systems (Linux/Unix, Windows 7 or better and Mac OS/OS X 10.6 or newer) together with a number of smaller platforms.

PROGRAMMING LANGUAGE

Python 2.7.14, 3.5.5 and 3.6.6.

ADDITIONAL SYSTEM REQUIREMENTS

This software has no particularly unusual system requirements and should operate on any typical desktop machine produced in the last five years. The calculations are CPU bound so the greater the single thread performance of the machine the more swiftly the computations will complete.

DEPENDENCIES

bayesint uses the following Python packages which are installed when *bayesint* is installed:

- *SymPy* ≥ 1.1.1 [14] A library for symbolic mathematics used to calculate the densities and distributions used in our credible intervals.
- *SciPy* ≥ 0.19.1 [9]. A library used for scientific computing used to obtain a χ^2 distribution for testing and to minimise the functions considered in the credible intervals.
- *NumPy* ≥ 1.13.3 [16]. A library with high-level mathematical functions used for vectorising our *SymPy* function prior to minimisation.
- *mpmath* ≥ 0.19 [8]. A library for arbitrary floating point arithmetic used in the minimisation. It is loaded through *SymPy*.

SOFTWARE LOCATION

Archive

Name: PyPI
Persistent identifier: *pypi.org/project/bayesint*
Licence: Open Government Licence 3.0
Publisher: Public Health England
Version published: 1.0.3
Date published: 11/04/2019

Code repository

Name: GitHub
Persistent identifier: *github.com/PublicHealthEngland/bayesint*
Licence: Open Government Licence 3.0
Date published: 11/04/2019

LANGUAGE

English

(3) REUSE POTENTIAL

We broadly outline the reuse potential of this software in the discussion section of this work but as a recapitulation this package permits the calculation of credible intervals wherever 2×2 type contingency table data appear in science. With the additional benefit that prior knowledge may be incorporated into this estimate in a truly Bayesian manner. Issues and feedback about the software can be provided by users through the GitHub issue logging system.

FUNDING STATEMENT

The research leading to these results has received UK government grant-in-aid funding. MBND and TJRF were funded by the National Institute for Health Research—Health Protection Research Unit for Modelling Methodology (HPRU MM). The funding bodies (UK government and HPRU MM) did not play any role in the formulation of this manuscript. The views expressed in this publication are those of the authors and not necessarily those of Public Health England.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

M. Bekker-Nielsen Dunbar *orcid.org/0000-0002-7249-3524*
Emergency Response Department, Public Health England, GB

Thomas J. R. Finnie *orcid.org/0000-0001-5962-4211*
Emergency Response Department, Public Health England, GB

REFERENCES

1. **Bekker-Nielsen Dunbar M, Finnie TJR, Sloane B, Hall IM.** Methods for calculating credible intervals for ratios of beta distributions with application to relative risks of premature death from the second plague pandemic. *PLOS ONE*. 2019; 14(2): e0211633. DOI: *https://doi.org/10.1371/journal.pone.0211633*
2. **Bentivoglio M, Pacini P.** Filippo Pacini: A determined observer. *Brain Res. Bull.* 1995; 38(2): 161–65. DOI: *https://doi.org/10.1016/0361-9230(95)00083-Q*
3. **Chen M-H, Shao Q-M.** Monte Carlo estimation of Bayesian credible and HPD intervals. *J. Comput. Graph. Stat.* 1999; 8(1): 69–92. DOI: *https://doi.org/10.1080/10618600.1999.10474802*
4. **Greenwood M, Yule GU.** The statistics of anti-typhoid and anti-cholera inoculations, and the interpretation of such statistics in genera. *Proc. R. Soc. Med.* 1915; 8(Sect. Epidemiol. State Med.): 113–194. DOI: *https://doi.org/10.1177/003591571500801433*
5. **Haffkine WM.** *Protective Inoculation Against Cholera*. Calcutta: Thacker, Spink & Co; 1913.
6. **Hedges LV, Gurevitch J, Curtis PS.** The meta-analysis of response ratios in experimental ecology. *Ecology*. 1999; 80(4): 1150–1156. DOI: *https://doi.org/10.1890/0012-9658(1999)080[1150:TMAORR]2.0.CO;2*
7. **Howard-Jones N.** Robert Koch and the cholera vibrio: A centenary. *BMJ (Clin. Res. Ed.)*. 1984; 288(6414): 379–81. DOI: *https://doi.org/10.1136/bmj.288.6414.379*
8. **Johansson F,** et al. *mpmath: A Python Library For Arbitrary-Precision Floating-Point Arithmetic*; 2013. URL *http://mpmath.org*.
9. **Jones E, Oliphant T, Peterson P,** et al. *SciPy*: Open source scientific tools for Python; 2001. URL *http://www.scipy.org*.
10. **Kastenbaum MA, Bowman KO.** Tables for determining the statistical significance of mutation frequencies. *Mutat. Res.* 1970; 9(5): 527–549. DOI: *https://doi.org/10.1016/0027-5107(70)90038-2*
11. **Lajeunesse MJ.** On the meta-analysis of response ratios for studies with correlated and multi-group designs. *Ecology*. 2011; 92(11): 2049–2055. DOI: *https://doi.org/10.1890/11-0423.1*
12. **Lajeunesse MJ.** Bias and correction for the log response ratio in ecological meta-analysis. *Ecology*. 2015; 96(8): 2056–2063. DOI: *https://doi.org/10.1890/14-2402.1*
13. **Matsen IV FA, Small CT, Soliven K, Engel GA, Feeroz MM, Wang X, Craig KL, Hasan MK, Emerman M, Linial ML, Jones-Engel L.** A novel Bayesian method for detection of apobec3-mediated hypermutation and its application to zoonotic transmission of simian foamy viruses. *PLOS Comput. Biol.* 02 2014; 10(2): 1–14. DOI: *https://doi.org/10.1371/journal.pcbi.1003493*
14. **Meurer A, Smith CP, Paprocki M, Čertik O, Kirpichev SB, Rocklin M, Kumar A, Ivanov S, Moore JK, Singh S, Rathnayake T, Vig S, Granger BE, Muller RP, Bonazzi F, Gupta H, Vats S, Johansson F, Pedregosa F, Curry MJ, Terrel AR, Roučka Š, Saboo A, Fernando I, Kulal S, Cimrman R, Scopatz A.** *SymPy*: Symbolic computing in Python. *PeerJ Comput. Sci.* 2017; 3: e103. DOI: *https://doi.org/10.7717/peerj-cs.103*
15. **Nurminen M, Mutanen P.** Exact Bayesian analysis of two proportions. *Scand. J. Stat.* 1987; 14(1): 67–77.
16. **Oliphant TE.** *Guide To NumPy*. CreateSpace Independent Publishing Platform, 2nd edition; 2015.
17. **Pham-Gia T.** Distributions of the ratios of independent beta variables and applications. *Commun. Stat. – Theory Methods*. 2000; 29(12): 2693–2715. DOI: *https://doi.org/10.1080/03610920008832632*
18. **Sherman IW.** *The Power Of Plagues*. ASM Press, 2nd edition; 2017. DOI: *https://doi.org/10.1128/9781683670018*
19. **Sverdlov O, Ryeznik Y, Wu S.** Exact Bayesian inference comparing binomial proportions, with application to proof-of-concept clinical trials. *Ther. Innov. Regul. Sci.* 2015; 49(1): 163–174. DOI: *https://doi.org/10.1177/2168479014547420*
20. **Python Software Foundation.** *Python Software, Version 2.7*, 2010. URL *http://www.python.org*.

TO CITE THIS ARTICLE:
Bekker-Nielsen Dunbar M, Finnie TJR 2021 bayesint: A Python Package for Calculating Bayesian Credible Intervals of Ratios of Beta Distributions. *Journal of Open Research Software*, 9: 35. DOI: *https://doi.org/10.5334/jors.283*

Submitted: 19 June 2019 Accepted: 11 November 2021 Published: 21 December 2021

Journal of Open Research Software is a peer-reviewed open access journal published by Ubiquity Press.