SoftwareX 6 (2017) 36–41

Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx

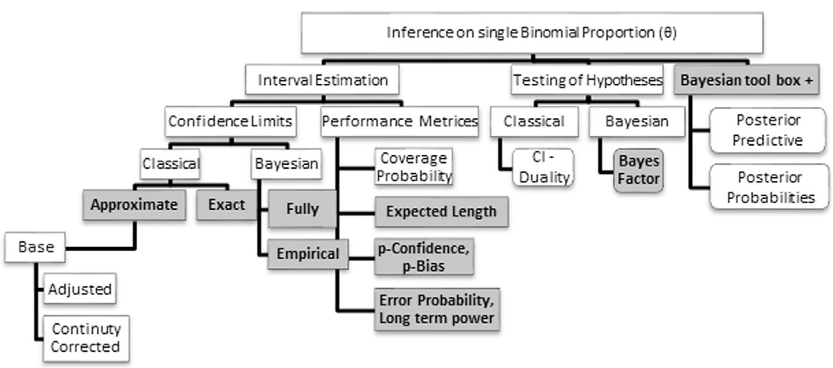
proportion: A comprehensive R package for inference on single Binomial proportion and Bayesian computations

M. Subbiah [a,*], V. Rajeswaran [b]

[a] Department of Mathematics, L N Government College, Ponneri, Tamil Nadu, India

[b] HCL Technologies Ltd., Chennai, Tamil Nadu, India

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 26 January 2016
Received in revised form
13 June 2016
Accepted 5 January 2017

Keywords:
Binomial proportion
Evaluation criteria
Predictive distributions
R package

ABSTRACT

Extensive statistical practice has shown the importance and relevance of the inferential problem of estimating probability parameters in a binomial experiment; especially on the issues of competing intervals from frequentist, Bayesian, and Bootstrap approaches. The package written in the free R environment and presented in this paper tries to take care of the issues just highlighted, by pooling a number of widely available and well-performing methods and apporting on them essential variations. A wide range of functions helps users with differing skills to estimate, evaluate, summarize, numerically and graphically, various measures adopting either the frequentist or the Bayesian paradigm.

Code metadata

Current Code version	*1.2.0*
Permanent link to code/repository used of this code version	https://github.com/ElsevierSoftwareX/SOFTX-D-16-00020
Legal Code License	*GPL-2* http://www.gnu.org/licenses/old-licenses/gpl-2.0.en.html
Code Versioning system used	*git*
Software Code Language used	*R*
Compilation requirements, Operating environments & dependencies	*Linux, OS X, Microsoft Windows. Runs within the R software environment.*
If available Link to developer documentation/manual	https://cran.r-project.org/web/packages/proportion/proportion.pdf
Support email for questions	v.rajeswaran@gmail.com

* Correspondence to: Department of Mathematics, L N Government College, Ponneri–601 204, Tamil Nadu, India.

E-mail address: sisufive@gmail.com (M. Subbiah).

http://dx.doi.org/10.1016/j.softx.2017.01.001
2352-7110/

Software metadata

Current software version	*1.2.0*
Permanent link to executables of this version	https://github.com/RajeswaranV/proportion
Legal Software License	*GPL-2* http://www.gnu.org/licenses/old-licenses/gpl-2.0.en.html
Computing platform/Operating System	*Linux, OS X, Microsoft Windows*
Installation requirements & dependencies	*R 3.2.0 or above from* https://cran.r-project.org/
If available Link to user manual - if formally published include a reference to the publication in the reference list	https://cran.r-project.org/web/packages/proportion/proportion.pdf
Support email for questions	v.rajeswaran@gmail.com

1. Introduction

Statistical inference for binomial proportion (θ) from a sequence of n independent Bernoulli trial has witnessed active research discussions and has found wider applications; precisely inferential procedures that deal with confidence intervals (CI) are quite pervasive in frequentist (approximate or exact) and Bayesian cultures (including constructing subjective or informative priors) as well as bootstrap approach. Methodological and practical problems on different domains such as bio-statistics, epidemiology, and ecology can be found from an extensive literature on comparison of different procedures, evaluation criterions such as coverage probability (CP), expected length (EL) and applications; [1–14] provides few imperative studies (since 2002) with regard to this century old yet active and interesting statistical problem.

Especially, the problem has become much more interesting to investigate in terms of constructing new methods and/or evaluating the performance of different methods for sparse data or behavior of θ in its boundary. To alleviate such issues, transformations (logit, arcsine), addition of pseudo quantities or continuity correction (CC) is employed with basic approximation methods; on the other hand, exact methods with various modifications to improve the performance and conjugate hierarchical Bayesian (HB) method using non-informative prior can also be found in various discussions. Additionally, computing facilities has also drawn considerable attention in a very specific manner [3,15–17] in that approximate, exact or Bayesian procedures have been discussed without comprehensively providing a computing tool.

In particular, a statistical programming environment R [18] has two built-in functions that are based only on inverting tests; first one, binom.test that performs an exact test of a simple null hypothesis about the probability of success in a Bernoulli experiment and Clopper–Pearson test with option for two or one sided, with default hypothesized probability of success as 0.5 and 0.95 level confidence interval is the default; second one is the prop.test for testing the equality of proportions (probabilities of success) in several groups. The latter function has a warning message regarding chi-square approximations for small $n(<10)$ and follows only Yates' continuity correction; alternately, prop.test includes Wilson Score method with CC except when x is near $n/2$ and this limitation has some undesirable effect as observed in Pires [3].

Further, few packages in R provide similar and more functions; following table in alphabetical order provides a quick summary of widely perceptible packages but may not be exhaustive.

Name	Details
BayesFactor	Bayes testing for simple hypothesis only with logistic prior
binom	Exact - Only Clopper–Pearson as binom.test, Asymptotic Wald Interval, Agresti–Coull method, Wilson method (without CC); transformed Wald approach, based on logit complementary log, probit function, Likelihood ratio method, and Profile likelihood. Bayes, conjugate beta with Jeffreys as default and one sided intervals for boundary values. Other functions provide CPs, ELs, root mean squared values, plots for CP.
binomSamSize	Mainly for sample size determination but has two exact and approximate methods
BlakerCI	Specifically for Blaker exact method
epiR	Score CI and Bayes with beta parameters from expert opinion
prevalence	Wald, Agresti–Coull, Clopper–Pearson, Jeffreys, Score interval
prop.comb.RR	A more general for measures using θ but can be used for single proportion that gives CIs based on Score (with and without CC), Adjusted Arc Sine, Adjusted Wald, Modified Score and Exact (Clopper–Pearson)
PropCIs	Score, Agresti–Coull (two options), Blaker, Clopper–Pearson, Mid-p

However, requirements generally differ among many studies in terms of the choice of base methods and their alternatives, evaluation criteria, simpler forms, and teaching purpose. This work has envisaged a need and scope for an all-inclusive package to include existing procedures of extensive usage, expand the scope of generalization, designing a comparative mechanism both numerically and graphically, and include more Bayesian procedures.

Hence it can be noticed that, though some software (open source/commercial) accommodate this inferential problem, following aspects need further attention; (i) inclusion of many estimation procedures and their derived variants in more generalized form, (ii) appropriate tools to evaluate the procedures using different parametric values, (iii) Bayesian computing procedures beyond confidence intervals such as Bayes factor for hypotheses testing, predictive inferences, posterior probabilities and Empirical Bayes (EB) procedures, and (iv) suitable graphical presentations to compare different procedures for their performances.

The wide spread usage in terms of research, practice, and text books inclusion are the sole criteria to choose eight approximate procedures of constructing CI for θ; suitable methods are considered for adjusting the data, n and number of successes $X \sim \text{Binomial}(n, \theta)$ or adopting continuity corrections. Such variants may not be required for exact and Bayesian methods though studies like [19] recommend boundary adjusted Bayesian intervals. Also, most visible Clopper–Pearson (exact) method has a computing flexibility using beta function or F-distribution; however, a general way to obtain CI due to exact method has been incorporated in the present work.

In a similar spirit, CP and EL have been included based on their extensive presence in the research and practical studies;