CSIRO PUBLISHING

Publications of the Astronomical Society of Australia, 2011, **28**, 128–139 www.publish.csiro.au/journals/pasa

On the Estimation of Confidence Intervals for Binomial Population Proportions in Astronomy: The Simplicity and Superiority of the Bayesian Approach

Ewan Cameron

Department of Physics, Swiss Federal Institute of Technology (ETH Zurich), CH-8093 Zurich, Switzerland. Email: cameron@phys.ethz.ch

Received 2010 December 3, accepted 2011 March 1

Abstract: I present a critical review of techniques for estimating confidence intervals on binomial population proportions inferred from success counts in small to intermediate samples. Population proportions arise frequently as quantities of interest in astronomical research; for instance, in studies aiming to constrain the bar fraction, active galactic nucleus fraction, supermassive black hole fraction, merger fraction, or red sequence fraction from counts of galaxies exhibiting distinct morphological features or stellar populations. However, two of the most widely-used techniques for estimating binomial confidence intervals — the 'normal approximation' and the Clopper & Pearson approach — are liable to misrepresent the degree of statistical uncertainty present under sampling conditions routinely encountered in astronomical surveys, leading to an ineffective use of the experimental data (and, worse, an inefficient use of the resources expended in obtaining that data). Hence, I provide here an overview of the fundamentals of binomial statistics with two principal aims: (I) to reveal the ease with which (Bayesian) binomial confidence intervals with more satisfactory behaviour may be estimated from the quantiles of the beta distribution using modern mathematical software packages (e.g. R, MATLAB, MATHEMATICA, IDL, PYTHON); and (II) to demonstrate convincingly the major flaws of both the 'normal approximation' and the Clopper & Pearson approach for error estimation.

Keywords: methods: data analysis — methods: statistical

1 Introduction

One problem frequently encountered in astronomical research is that of estimating a confidence interval (CI) on the value of an unknown population proportion based on the observed number of success counts in a given sample. The unknown population proportion may be, for instance, the intrinsic fraction of barred disk galaxies at a specific epoch to be inferred from the observed number of barred disks in a volume-limited sample (e.g., Elmegreen et al. 1990; van den Bergh 2002; Cameron et al. 2010; Nair & Abraham 2010), with the corresponding binomial CI used to evaluate the hypothesis that the bar fraction changes with redshift relative to a local benchmark (e.g., Cameron et al. 2010). Experiments to investigate the role of mass and environment in quenching star-formation via measurement of the galaxy red sequence fraction (e.g., Baldry et al. 2006; Hester et al. 2010; Ilbert et al. 2010), or to investigate whether or not major mergers were more frequent at high redshift via measurement of the close-pair/asymmetric fraction (e.g., De Propris et al. 2005; Conselice et al. 2008; López-Sanjuan et al. 2010), also routinely present this class of problem.

However, the two most commonly used methods for estimating CIs on binomial population proportions — the 'normal approximation' and the Clopper & Pearson (1934) approach — exhibit significant flaws under routine sampling conditions (cf. Vollset 1993; Santner 1998; Brown et al. 2001, 2002). In particular, the 'normal approximation' (also called the 'Poisson error') may systematically underestimate the CI width necessary to provide coverage at the desired level, especially for small samples, but even for rather large samples when the true population proportion is either very low or very high. If used naïvely the 'normal approximation' has the potential to mislead one into over-stating the significance of one's inferences concerning the physical system under study formulated on the basis of the observed data.

Astronomers aware of these flaws in the 'normal approximation' often adopt the alternative Clopper & Pearson (1934) approach to CI estimation by way of reference to the CI tables in Gehrels (1986). Unfortunately, the Clopper & Pearson (1934) approach suffers from the opposite problem to that of the 'normal approximation' — namely, a systematic over-estimation of the CI width required to provide the desired coverage (Clopper & Pearson 1934; Neyman 1935; Gehrels 1986; Agresti & Coull 1998). In scientific research this over-estimation of the statistical measurement uncertainties may mislead one into placing insufficient confidence in the experimental outcomes, resulting in an inefficient use of the measured data (and hence the resources expended in obtaining that data). Indeed, it has been well argued by Agresti & Coull

 10.1071/AS10046 1323-3580/11/02128

(1998) that in many practical applications even the 'normal approximation', despite its flaws, is preferable to the Clopper & Pearson (1934) approach.

Fortunately, there exist a multitude of alternative methods for generating CIs on binomial population proportions, many of which exhibit far more satisfactory behaviour than either the 'normal approximation' or the Clopper & Pearson (1934) approach — see Agresti & Coull (1998) and Brown et al. (2001) for various examples. Here I review both the theory and application of one of these methods — use of the beta distribution quantiles — deriving from a simple Bayesian analysis in which a uniform ('non-informative') prior is adopted for the true population proportion (e.g., Gelman et al. 2003). As I will demonstrate, the beta distribution generator for binomial CIs is both theoretically well motivated and easily applied in practice using widely available mathematical software packages (e.g., R, MATLAB, MATHEMATICA, IDL, PYTHON). Ultimately, I advocate strongly that this strategy for estimating binomial CIs be adopted in future studies aiming to constrain fundamental population proportions in astronomical research (e.g., the galaxy bar fraction, red sequence fraction, or merger fraction) — especially for samples intrinsically of small to intermediate size, or when the subdivision of larger samples for analytical purposes produces sparsely populated data bins.

2 The Binomial Distribution

In probability theory any experiment for which there are only two possible random outcomes — *success*, occurring with probability p, or *failure*, occurring with probability $q = (1 - p)$ — is referred to as a *Bernoulli trial*. Examples of Bernoulli trials in astronomical research may include asking whether or not a randomly sampled galaxy is barred, red-sequence, or merging. The probability, P, of observing a particular number of successes, k, in a series of n independent Bernoulli trials (with common success probability p) is governed by the *binomial* probability function:[1]

$$P(k, n, p) = \binom{n}{k} p^k q^{n-k} \tag{1}$$

where $0 \leq k \leq n$, $k \in \mathbb{Z}$ (an integer), and

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

(see, for example, Quirin 1978). Note that the probabilities given by the $n + 1$ possible values of k correspond to the $n + 1$ terms of the binomial expansion of $(p + q)^n$. The number of barred systems counted in a given sample of disk galaxies is a classic example of a binomially distributed variable in astronomy. The corresponding expectation value for the number of successes is $\sum_{k=0}^{n} k \times P(k, n, p) = np$ with a variance of $\sum_{k=0}^{n} (k - np)^2 \times P(k, n, p) = npq$. Moreover, the expectation value for the *fraction* of successes, k/n, is equal to the Bernoulli trial success probability (also referred to as the 'underlying population proportion'), p, and its variance is pq/n.

2.1 An Intermission: Just What Is a Confidence Interval?

As explained eloquently by both Kraft et al. (1991) and Ross (2003), there is a fundamental difference between the 'classical' and 'Bayesian' definitions of the term 'confidence interval'. In classical statistical theory a binomial CI is defined as a pair of random variables, P_l and P_u, (with each random variable necessarily a finite, real-valued, measurable function; cf. Rao & Swift 2006) operating on the set of all possible experimental outcomes, $\theta = \{k: 0 \leq k \leq n, k \in \mathbb{Z}\}$, such that if the experiment were to be repeated by a sufficiently large number of independent observers then the *fraction of observers* for whom the true value of the underlying population proportion is covered by their realisation of these random variables — i.e., for whom $P_l(\theta_i) = p_l < p < p_u = P_u(\theta_i)$ — is guaranteed to converge to (at least) a specific value, c, termed 'the confidence level'. In the Bayesian paradigm, on the other hand, the underlying population proportion is treated as an unknown model parameter and the binomial CI defined as an interval, (p_l, p_u), to which the experimenter believes may be assigned a *probability*, c, of containing the true value of p, based upon consideration of the likelihood function for p given the experimental data and the strength of any *a priori* beliefs or expectations regarding the system under study. (Indeed, acknowledging the significant conceptual differences between these alternative approaches to the binomial CI, the term 'credible interval' is often used instead in Bayesian analysis to avoid confusion with the classical nomenclature.) Importantly, as noted by Kraft et al. (1991), regardless of one's philosophical position regarding these two statistical systems, 'the Bayesian definition of confidence intervals reflects common astronomical usage better than the classical definition'.

Of the three binomial CI generators discussed in this review, only that attributed to Clopper & Pearson (1934) is consistent with the classical definition for all possible values of the underlying population proportion and sample size. However, I will argue that (at least) in the case of the binomial distribution, Bayesian CIs provide generally more satisfactory behaviour for astronomical purposes than their classical counterparts, even when evaluated against a performance diagnostic based on the classical definition — namely, the coverage fraction (or 'effective coverage') at given p and n.

[1] One may note that the correct terminology in a statistical context is actually 'binomial probability *mass* function', owing to the discrete nature of the binomial distribution, i.e., that there exist a finite number of possible k values (the integers from 0 to n, inclusive) to which non-zero probabilities may be assigned. (As distinct from the alternative case of a 'probability *density* function', such as the Bayesian posterior probability distribution for p considered in Section 3, for which non-zero probabilities may only be assigned to measurable intervals on the real number line, and not individual — or even countable sets of — real numbers.) Nevertheless, to avoid any confusion with the more commonly used definition of the term 'mass function' in astronomy I adopt the shorter expression 'binomial probability function' herein.

https://doi.org/10.1071/AS10046 Published online by Cambridge University Press

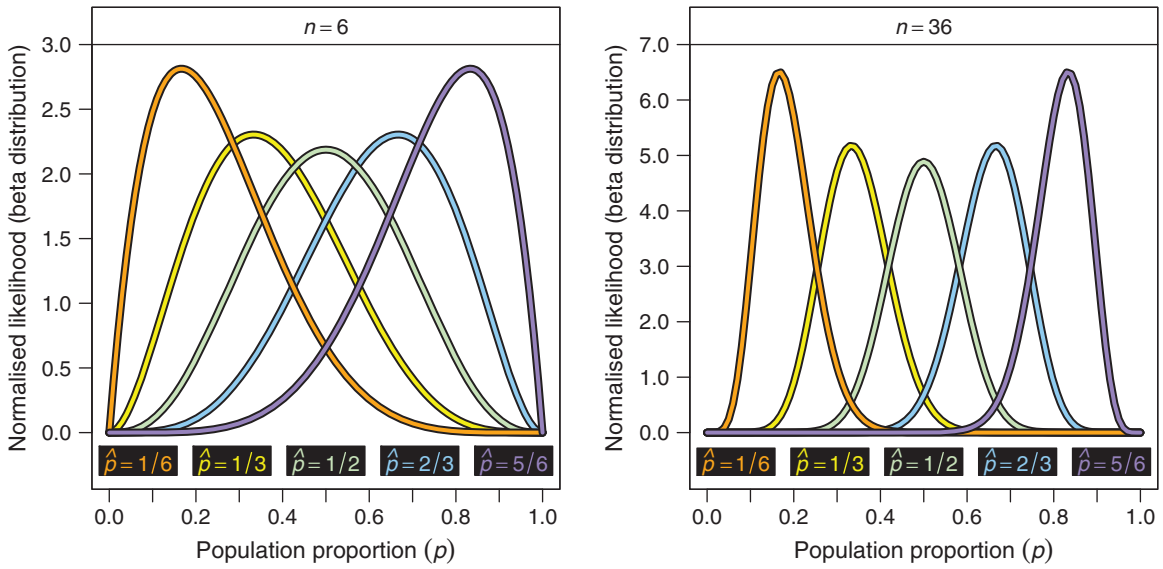

Figure 1 Example likelihood functions for the true value of the underlying population proportion, p, given five 'measured' success fractions, $\hat{p} = k/n$, for samples of sizes $n = 6$ (left panel) and $n = 36$ (right panel). In each case the shape of the curve is given by the beta distribution with shape parameters as specified by Eqn 2. The asymmetric nature of this likelihood function in the small sample size regime is clearly evident among the $n = 6$ examples, as is its convergence in the intermediate to large sample size regime towards a narrower, more symmetric, (pseudo-)normal distribution among the $n = 36$ examples.

3 The Beta Distribution Generator for Binomial Confidence Intervals

In astronomical data analysis it is standard practice to adopt the measured success fraction (also referred to as the 'observed population proportion'), $\hat{p} = k/n$, as one's 'best guess' of the underlying population proportion. In statistical terms, $\hat{p}$ is employed as a *point estimator* for p. The likelihood of observing the result, $\hat{p} = k/n$, for a given value of p is, of course, proportional to $p^k q^{n-k}$. Normalisation of this likelihood function over $0 < p < 1$ defines a 'beta distribution' with integer parameters $a = k + 1$ and $b = n - k + 1$:

$$B(a,b) = \frac{(a+b-1)!}{(a-1)!(b-1)!} p^{a-1} q^{b-1} \qquad (2)$$

where $q = 1 - p$ (e.g., Gelman et al. 2003; Ross 2003). Differentiation of this likelihood function reveals that our best guess, $\hat{p}$, is in fact the maximum likelihood estimator of p.[2] The characteristic shape of the (beta distribution) likelihood function for p is illustrated in Figure 1 at a variety of 'measured' success fractions for samples of sizes $n = 6$ (left panel) and $n = 36$ (right panel). At small n, the likelihood function for p is markedly asymmetric (except where $\hat{p} = 1/2$), but at intermediate n it is visibly converging towards a narrow, symmetric, (pseudo-)normal distribution — the motivation behind the 'normal approximation' discussed in Section 4.

Given no a priori knowledge to inform one's expectations regarding the experimental outcome, one may suppose that all values of p are equally 'probable'. Formally, this condition is characterised via the Bayes–Laplace uniform prior, for which $P_{\rm prior}(p) = 1$ over $0 < p < 1$. Application of Bayes' theorem under this assumption allows one to treat the normalised likelihood function for p as a posterior probability distribution. Thus, the quantiles of the beta distribution from Eqn 2 may be used directly to estimate (Bayesian) confidence intervals on the underlying population proportion given the observed data.[3] Specifically, the lower and upper bounds, p_l and p_u, defining an 'equal-tailed' (or 'central') interval for p at a nominal confidence level of $c = 1 - \alpha$ are given by the quantiles:

$$\int_0^{p_l} B(a,b)dp = \alpha/2 \quad \text{and} \quad \int_{p_u}^1 B(a,b)dp = \alpha/2. \qquad (3)$$

Note that the bounds of this 'equal-tailed' interval (which partition the probability of p greater than p_u equal to that of p less than p_l) will be necessarily asymmetric about the maximum likelihood value, $\hat{p}$, (except at $\hat{p} = 1/2$) owing to the asymmetric nature of the (beta distribution) likelihood function for p (shown in Figure 1). As I will demonstrate below, binomial CIs generated in this manner

[2] Technically, when $\hat{p} = 0$ (or 1) the likelihood function for p has no zero first derivative on the open interval, (0, 1), although the function itself is indeed strictly increasing as $p \to 0$ (or 1). In this case one may choose to adopt the median (50% quantile) of the (beta distribution) likelihood function as one's best guess for p, or else to compute a 'one-sided' confidence interval bounding p instead. In either case, one proceeds using similar principles.

[3] Astronomers familiar with the work of Burgasser et al. (2003) on binarity in brown dwarfs may be familiar with the algorithm for recovering confidence intervals on p given in their Appendix, which is in fact equivalent to the Bayesian approach with uniform prior presented here (although Burgasser et al. 2003 make no explicit reference to either Bayes' theorem or the beta distribution).

https://doi.org/10.1071/AS10046 Published online by Cambridge University Press

have one rather desirable property, not shared by either the 'normal approximation' or the Clopper & Pearson (1934) approach — namely, their *mean* effective coverage is consistently very close to the nominal confidence level, even in the small sample size regime.

In the upper panel of Figure 2, I examine first the effective coverage, c_e, of 'equal-tailed' binomial CIs defined via the beta distribution for a range of population proportions and sample sizes ($0.025 \leq p \leq 0.975$ and $1 \leq n \leq 100$) at a nominal level of $c_n \approx 0.683$ (1σ) — with the effective coverage defined as the fraction of samples drawn from the binomial probability function with given p and n for which the corresponding realisation of the CI under investigation encompasses the true population proportion. Thus, the effective coverage fractions, c_e, presented here are computed as the sum of all binomial probabilities $P(k, n, p)$ over $\{k : 0 \leq k \leq n,\ k \in \mathbb{Z}\}$, for which the triad $\{k, n, p\}$ produces a confidence interval, (p_l, p_u), containing (covering) p.

One of the most striking features of this plot is the remarkable sensitivity of the effective coverage to the true underlying population proportion and sample size. This so-called 'oscillation signature' is an inherent property of *all* deterministic (i.e., non-randomising) generators for binomial CIs, arising from the discreteness of the binomial distribution.[4] Despite these oscillations it is clear that the beta distribution CIs do achieve an effective coverage close to (or slightly greater than) the desired confidence level over the vast majority of the parameter space explored here. Indeed, even at the extremes of $p \lesssim 1/6$ and $p \gtrsim 5/6$, where the oscillations are initially rather large, there is evidently a rapid increase in coverage stability with increasing sample size, such that the 'oscillation signature' is vastly suppressed by $n \gtrsim 40$, and effectively eliminated (at least for $0.025 \leq p \leq 0.975$) by $n \gtrsim 80$ (unlike in the case of the 'normal approximation' examined in Section 4).

In the lower panel of Figure 2, I examine the corresponding *mean* effective coverage (averaged uniformly over $0.025 \leq p \leq 0.975$) as a function of sample size. Whereas the effective coverage at given p and n shown in the upper panel is consistent with the classical notion of confidence interval performance, the *mean* effective coverage may be considered a 'Bayesian' CI performance diagnostic — i.e., if one really does hold all p values equally 'probable' a priori, then one's favoured CI generator should be at least expected to provide coverage consistent with the nominal level in the long-term average of all equivalent experiments. Inspection of the lower panel of Figure 2 confirms a very close agreement between the mean effective coverage of the beta distribution CI generator and the nominal confidence level, independent of n.

Most modern mathematical software packages provide robust, easy-to-use library functions for computing beta distribution quantiles (e.g., the QBETA routine in R; the QUANTILE and BETADISTRIBUTION commands in MATHEMATICA; the BETAINCINV function in MATLAB; the IBETA function in IDL; or the DIST.BETA.PPF function in PYTHON). Explicit code fragments demonstrating the implementation of these commands are provided in the Appendix to this paper, and I advocate strongly the use of these recipes for the computation of confidence intervals on binomial population proportions in future astronomical studies. In Tables 1 and 2 in the Appendix, I present compilations of 'equal-tailed' CIs generated in this manner at nominal confidence levels of 1σ and 3σ, respectively, for all possible observed success counts in sample sizes up to $n = 20$. These tables are intended both as a convenient reference for use directly in studies involving samples of 20 objects or less, and as a benchmark against which to confirm the correct implementation of the beta distribution CI generator for users newly adopting this technique.

A note on the above The (non-informative) Bayes–Laplace uniform prior may in fact be viewed as the special case of $P_{\rm prior}(p) = B(1, 1) = 1$ within a wider family of possible conjugate priors for the binomial population proportion based on the beta distribution. Another popular non-informative prior for p is the Jeffreys prior of $P_{\rm prior}(p) = B(1/2, 1/2)$[5] (cf. Brown et al. 2001; Gelman et al. 2003), which is, by design, proportional to the square root of the Fisher information. Application of the Jeffreys prior returns a posterior probability distribution for p of $B(k + 1/2, n - k + 1/2)$. The *performance* of binomial CIs generated via beta distribution quantiles based on the Jeffreys prior differ insignificantly from those based on the uniform prior over $0.025 \leq p \leq 0.975$ when $n \gtrsim 2$ — consistent with the description of both these priors as 'non-informative' (i.e., that even for small sample sizes the shape of the posterior probability distribution in both cases is strongly governed by the likelihood function of the observed data). (See the recent review by Cousins, Hymes & Tucker 2009 for a thorough evaluation of the performance of Bayesian CIs constructed with the Jeffreys prior.) Hence, while the specific results presented in this paper are computed exclusively using the uniform

[4] Brown et al. (2001) describe the 'oscillation signature' as the challenge of 'lucky p, lucky n' — namely that for certain ('lucky') combinations of underlying population proportion and sample size there exist two almost equally likely $\hat{p}$ values closely straddling the true p. For instance, if $p = 1/5$ and one has a sample of size $n = 3$, the possible $\hat{p}$ values are 0, 1/3, 2/3, and 1, occurring with frequencies 0.512, 0.384, 0.096, and 0.008, respectively. Tailoring a binomial CI specifically to this situation, one could define $p_l = \hat{p} - 2/15 - \epsilon$ and $p_u = \hat{p} + 1/5 + \epsilon$ (with ε an arbitrarily small constant necessary to ensure p is contained within the open interval (p_l, p_u) for $k = 0$ and 1), returning an effective coverage of $c_e = 0.512 + 0.384 = 0.896$. However, applying the same CI generator to a system with $p = 1/3$ (and again $n = 3$) for which the possible $\hat{p}$ values occur with frequencies 0.296, 0.444, 0.222, and 0.037 (rounded to 3 decimal places), one obtains an effective coverage of only $c_e = 0.444$! For further discussion of the impact of the 'oscillation signature' on binomial CIs the interested reader is referred to Agresti & Coull (1998) and Brown et al. (2001, 2002).

[5] Note that the factorial functions used in the beta distribution definition of Eqn 2 must be replaced by gamma functions according to the relation $(m)! = \Gamma(m + 1)$ in order to handle the non-integer input in this case.

https://doi.org/10.1071/AS10046 Published online by Cambridge University Press

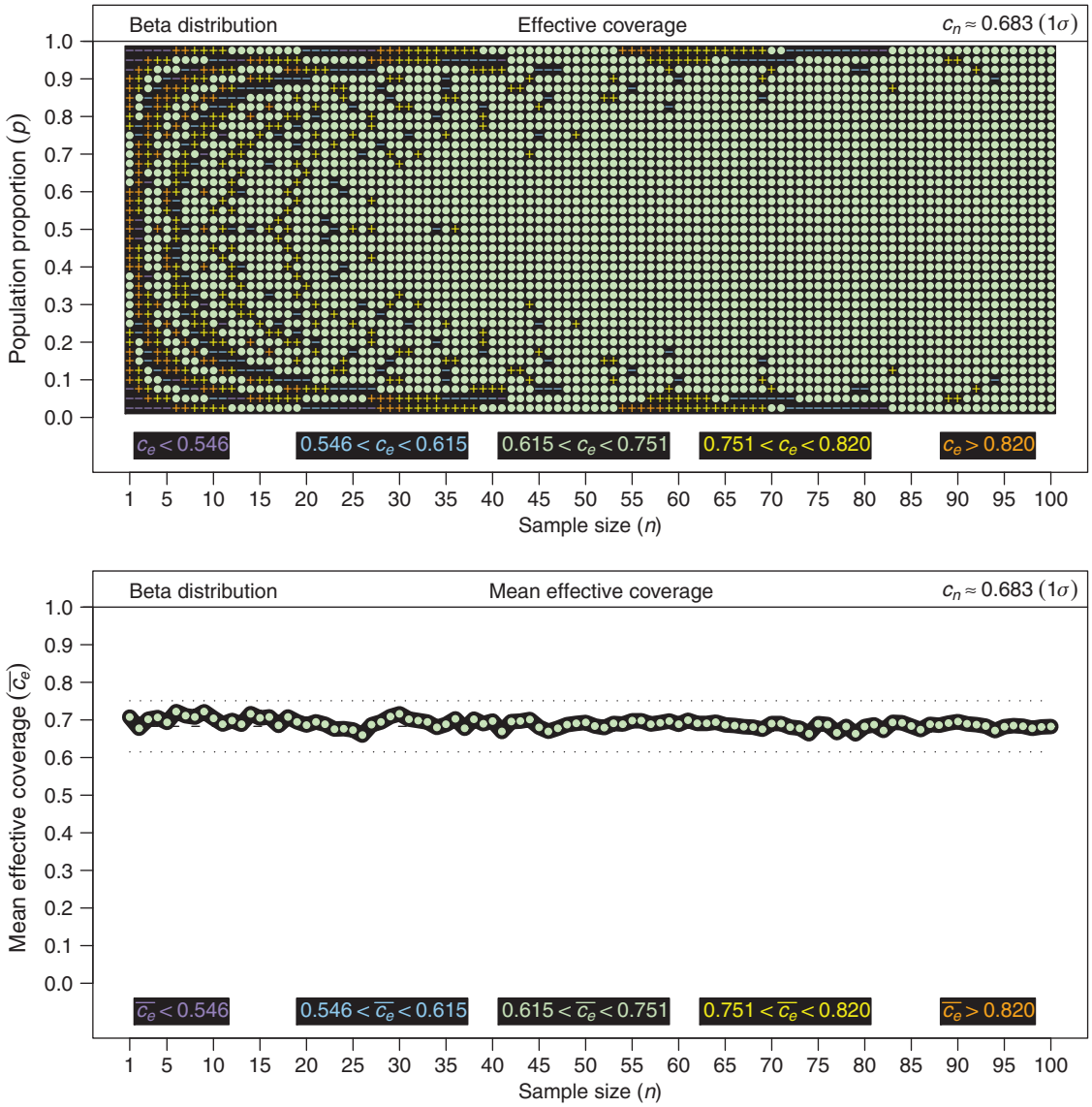

Figure 2 The effective coverage, c_e, of confidence intervals on the binomial population proportion generated from quantiles of the beta distribution at a nominal level of $c_n \approx 0.683\ (1\sigma)$ over the range $0.025 \leq p \leq 0.975$ and $1 \leq n \leq 100$ (upper panel). Averaging the measured c_e values uniformly over all p at each n returns the *mean* effective coverage as a function of sample size (lower panel).

prior, for the purposes of our general discussion regarding the superiority of the beta distribution quantile technique over the 'normal approximation' and the Clopper & Pearson (1934) approach, these two non-informative priors may be considered interchangable.

4 The 'Normal Approximation'

For a system with an underlying binomial population proportion, p, neither very close to 0 or 1, one may suppose (with reference to the central limit theorem) that the distribution of the $\hat{p}$ statistic in a series of independent samples of a fixed 'large' size will follow approximately a normal distribution. Under the assumptions of this 'normal approximation' (also called the 'Poisson error') one may employ the standard Wald test criterion, established by Wald & Wolfowitz (1939), to construct a two-sided confidence interval for p. Specifically, at a confidence level of $c = 1 - \alpha$ one may expect that the true value of p lies within the interval

$$\hat{p} - z_{1\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \qquad (4)$$

where $\hat{q} = 1 - \hat{p}$, and $z_{1-\alpha/2}$ is defined with reference to the standard normal distribution:

$$\int_{-\infty}^{z_{1-\alpha/2}} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)dx = 1 - \alpha/2.$$

Values of $z_{1-\alpha/2}$ for particular confidence levels may be obtained from reference tables in statistical textbooks (e.g., Quirin 1978) or computed within one's favourite mathematical software package (e.g., the QNORM function in R). Of course, the most commonly used formula for constructing error bars on measured galaxy bar fractions,

https://doi.org/10.1071/AS10046 Published online by Cambridge University Press

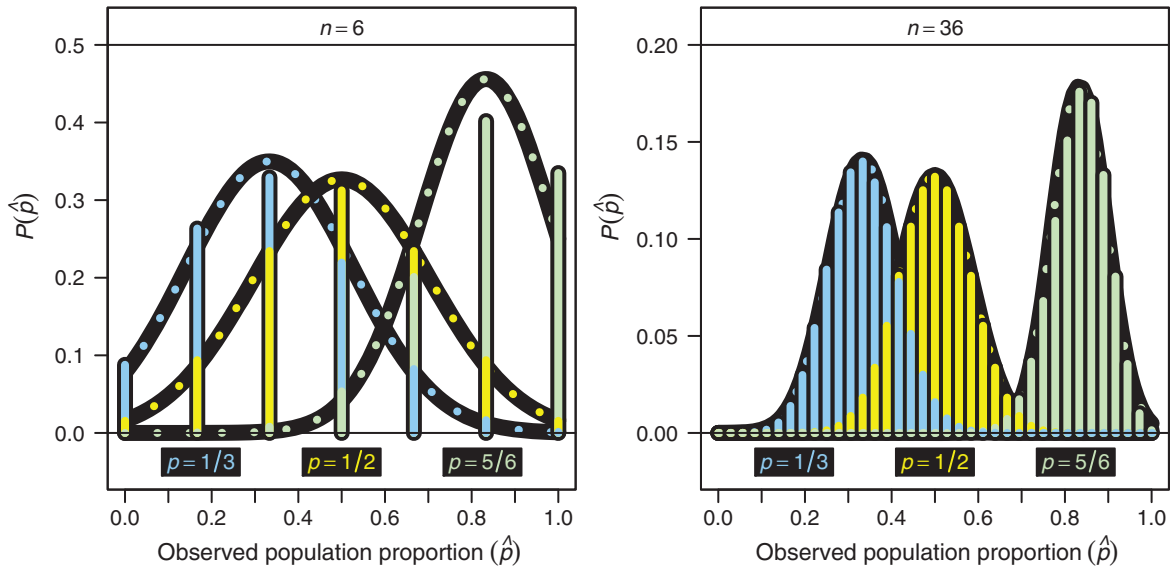

Figure 3 Comparison between the true binomial distribution of the $\hat{p}$ statistic (i.e., the observed population proportion k/n) and that assumed by the 'normal approximation'. Specifically, the $\hat{p}$ distributions of the binomial probability function with $p = 1/3$, 1/2, and 5/6 are contrasted against scaled normal distributions with matching means and variances for sample sizes of $n = 6$ (left panel) and $n = 36$ (right panel), respectively. In the small sample size regime the 'normal approximation' provides a reasonable representation of the $\hat{p}$ distribution at $p = 1/2$ and 1/3, but not 5/6, while in the intermediate to large sample size regime even the distribution at $p = 5/6$ is also clearly converging towards normal.

$p = \hat{p}\sqrt{\hat{p}\hat{q}/n}$ (e.g., Elmegreen et al. 1990), is simply the application of Eqn 4 at $z_{1-\alpha/2} = 1$, corresponding to a 1σ confidence level of $c \approx 0.683$. The cases of $z_{1-\alpha/2} = 2$ and 3 (i.e., 2σ and 3σ errors) correspond to higher confidence levels of $c \approx 0.954$ and 0.997, respectively.

As noted above, the key assumption behind this approach to binomial CI estimation — that the distribution of $\hat{p}$ may be approximated via a normal distribution with mean p and variance pq/n — is reasonable only under the conditions of a 'large' sample size and p neither very close to 0 or 1. In Figure 3 I compare the distribution of the $\hat{p}$ statistic (computed directly from the binomial probability function) against the shape of the corresponding 'normal approximation' for three different values of the underlying population proportion ($p = 1/3$, 1/2, and 5/6) and two different sample sizes ($n = 6$ and 36). In the small sample size example ($n = 6$) the 'normal approximation' provides a reasonable representation of the $\hat{p}$ distribution at $p = 1/3$ and $p = 1/2$, but performs poorly at $p = 5/6$ (i.e., p close to 1). However, in the intermediate sample size example ($n = 36$) there is now a clear convergence towards a normal distribution in $\hat{p}$ even at $p = 5/6$. These examples, presented in Figure 3, serve to illustrate the nature of deviations from 'normality' in the distribution of $\hat{p}$ at small n and/or extreme p values. The impact of these deviations on the performance of the 'normal approximation' as a binomial CI generator is examined below.

In Figure 4 I present the effective coverage of binomial CIs estimated via the 'normal approximation' as a function of p and n at a nominal confidence level of $c_n \approx 0.683$ (1σ). As in the case of the beta distribution quantile approach described above, there is a clear 'oscillation signature' visible in this figure, reflecting a marked sensitivity in the coverage performance to the value of the underlying population proportion and sample size.[6] However, it is also evident that the 'normal approximation' suffers a *systematic* decline in performance both for small n and towards extreme values of p near 0 or 1, generating binomial CIs with effective coverage far below the desired level. The strict *symmetry* of the 'normal approximation' CI about the observed success fraction — which at low or high $\hat{p}$ may even extend (unphysically) to $p \leq 0$ or $p \geq 1$ — regardless of the inherent *asymmetry* in the likelihood distribution for p (see Figure 1) is the principal cause of these coverage failures. The poor performance of the 'normal approximation' at small n is further highlighted in the corresponding plot of *mean* effective coverage against sample size shown in the lower panel of Figure 4. For the 1σ CIs examined here (and popularly adopted in studies of the galaxy bar fraction) the *mean* effective coverage of the 'normal approximation' is far below the nominal level for $n \lesssim 20$, and should thus be strictly avoided in this small

[6] It is important also to note that this 'oscillation signature' is evident even in binomial CIs generated via the 'normal approximation' *at very large sample sizes*, as thoroughly demonstrated by Brown et al. (2001, 2002). For instance, Brown et al. (2001) describe the erratic behaviour of the 'normal approximation' coverage at a nominal level of $c_n = 0.95$ for a system with $p = 0.005$, whereby there is a steady convergence in c_e towards 0.95 for n increasing until $n = 592$, at which point coverage falls suddenly to $c_e = 0.792$! Similarly, Brown et al. (2002) demonstrate that in order to ensure coverage stays at or above a nominal level of $c_n = 0.93$ for a system with $p = 0.1$ using the 'normal approximation' one requires a sample size of at least $n = 286$, whereas for the Bayesian (Jeffreys non-informative prior) CI this criterion is satisfied by $n = 47$.

https://doi.org/10.1071/AS10046 Published online by Cambridge University Press

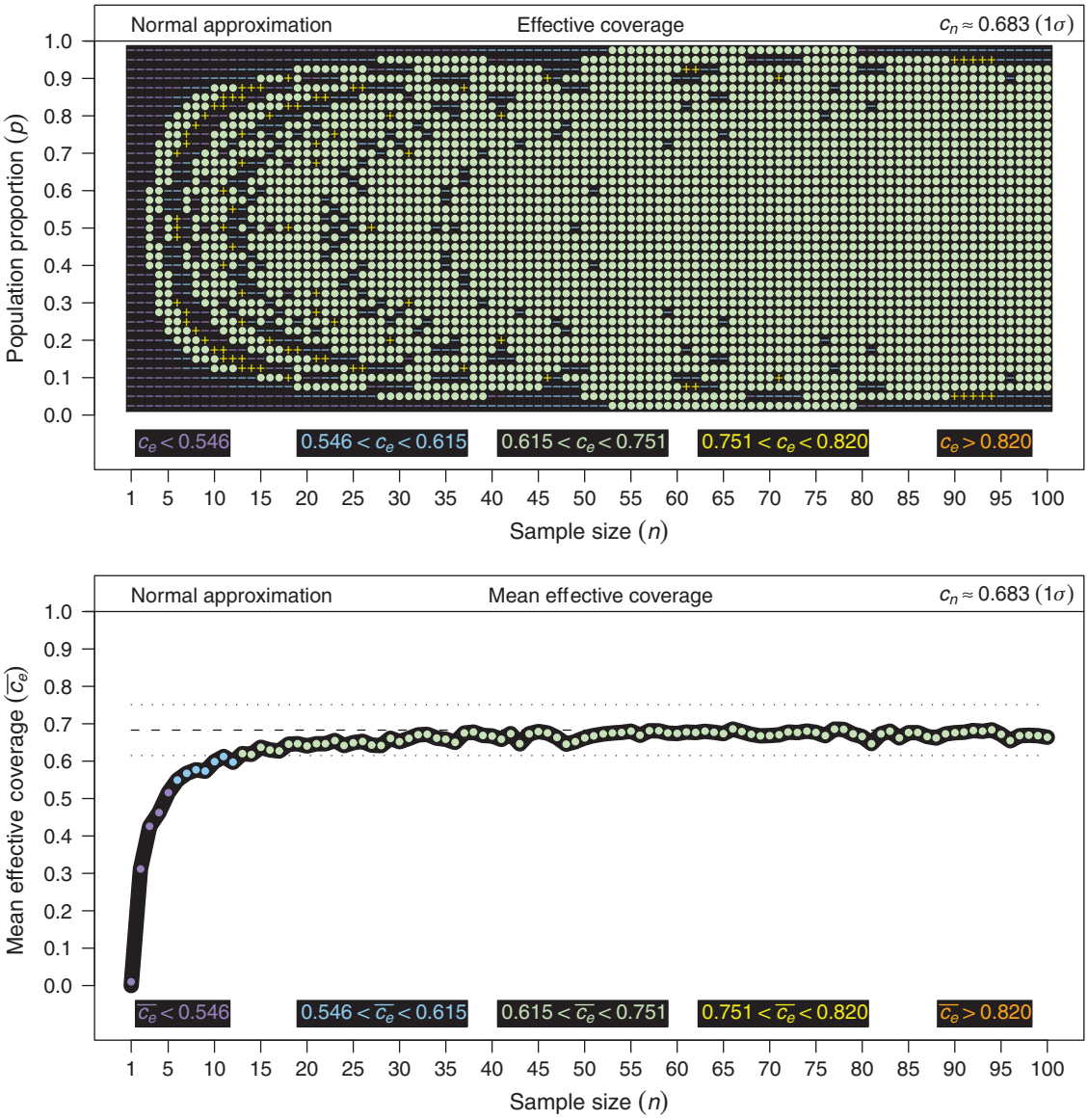

Figure 4 The effective coverage, c_e, of confidence intervals on the binomial population proportion generated via the 'normal approximation' at a nominal level of $c_n \approx 0.683\ (1\sigma)$ over the range $0.025 \leq p \leq 0.975$ and $1 \leq n \leq 100$ (upper panel). Averaging the measured c_e values unformly over all p at each n returns the *mean* effective coverage as a function of sample size (lower panel).

sample size regime. Indeed, although its *mean* effective coverage does ultimately improve with increasing n, one may be well advised to avoid the 'normal approximation' altogether in light of its poor effective coverage at extreme p values and the ready availability of a superior CI generator in the form of the (Bayesian) beta distribution quantiles described in Section 3.

The flaws in the 'normal approximation' as a CI generator described above were a great source of concern for statisticians in the 1930s, prompting the search for alternatives able to ensure universal coverage of at least the nominal level (thereby satisfying the classical definition of the term 'confidence interval') while remaining readily computable given the limited aids available at the time (such as reference tables of quantiles for standard distributions). The most popular of all such proposed alternatives was the Clopper & Pearson (1934) approach (cf. Gehrels 1986), which I review below.

5 The Clopper & Pearson Approach

In their landmark 1934 paper Clopper & Pearson presented a direct method for constructing 'classical' confidence intervals on inferred population proportions based on quantiles of the binomial probability function (Eqn 1), guaranteed to provide a coverage probability of at least (but likely exceeding) the nominal confidence level. The 'two-sided' (Clopper & Pearson 1934) CI at $c = 1 - \alpha$ is constructed by solving the following equations for the lower and upper bounds, $P_l(k) = p_l$ and $P_u(k) = p_u$:

$$\sum_{j=k}^{n} \binom{n}{j} p_l^j (1 - p_l)^{n-j} = \alpha/2 \ (\text{for } k \neq 0) \qquad (5)$$

https://doi.org/10.1071/AS10046 Published online by Cambridge University Press

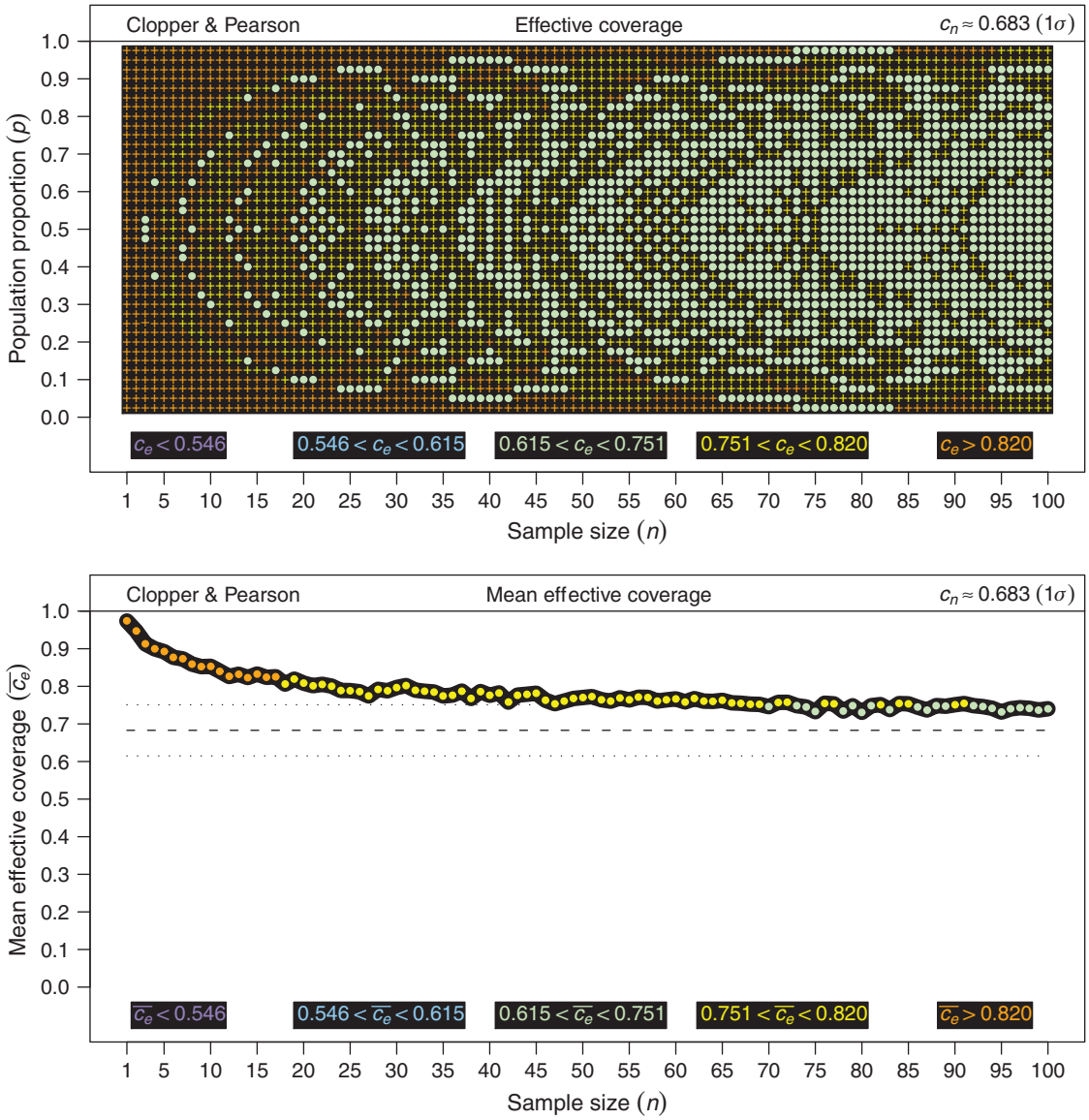

Figure 5 The effective coverage, c_e, of confidence intervals on the binomial population proportion generated via the Clopper & Pearson (1934) approach at a nominal level of $c_n \approx 0.683$ (1σ) over the range $0.025 \leq p \leq 0.975$ and $1 \leq n \leq 100$ (upper panel). Averaging the measured c_e values unformly over all p at each n returns the *mean* effective coverage as a function of sample size (lower panel).

and

$$\sum_{j=0}^{k} \binom{n}{j} p_u^j (1-p_u)^{n-j} = \alpha/2 \text{ (for } k \neq n) \qquad (6)$$

where k is again the observed number of successes (e.g., barred galaxies) in the sample, and n the total sample size. Note that in the extreme cases of $\hat{p} = 0$ or 1, the Clopper & Pearson (1934) formulae reduce simply to:

$$p_l = (\alpha/2)^{1/n} \text{ for } \hat{p} = 1 \text{ and} \qquad (7)$$

$$p_u = 1 - (\alpha/2)^{1/n} \text{ for } \hat{p} = 0. \qquad (8)$$

Modern mathematical software packages, such as R and MATLAB, support easy-to-use library functions (e.g., BINOM.TEST in the STATS package in R; or BINOFIT in the STATISTICS TOOLBOX in MATLAB) for computation of Clopper & Pearson (1934) confidence limits, which employ robust algorithms for the solution of Eqns 5 and 6. Alternatively, there exist numerous reference tables of pre-computed binomial CIs based on the Clopper & Pearson (1934) approach — most notably those of Gehrels (1986), a popular reference for estimating uncertainties in astronomical population proportions.

In the upper panel of Figure 5, I examine the effective coverage of CIs generated via the Clopper & Pearson (1934) approach as a function of p and n at a nominal confidence level of $c \approx 0.683$ (1σ). In contrast with the results for both the beta distribution and the 'normal approximation' reviewed above, the Clopper & Pearson (1934) CIs provide coverage far exceeding the nominal confidence level over much of this parameter space. The Clopper & Pearson (1934) coverage excess is also clearly

https://doi.org/10.1071/AS10046 Published online by Cambridge University Press

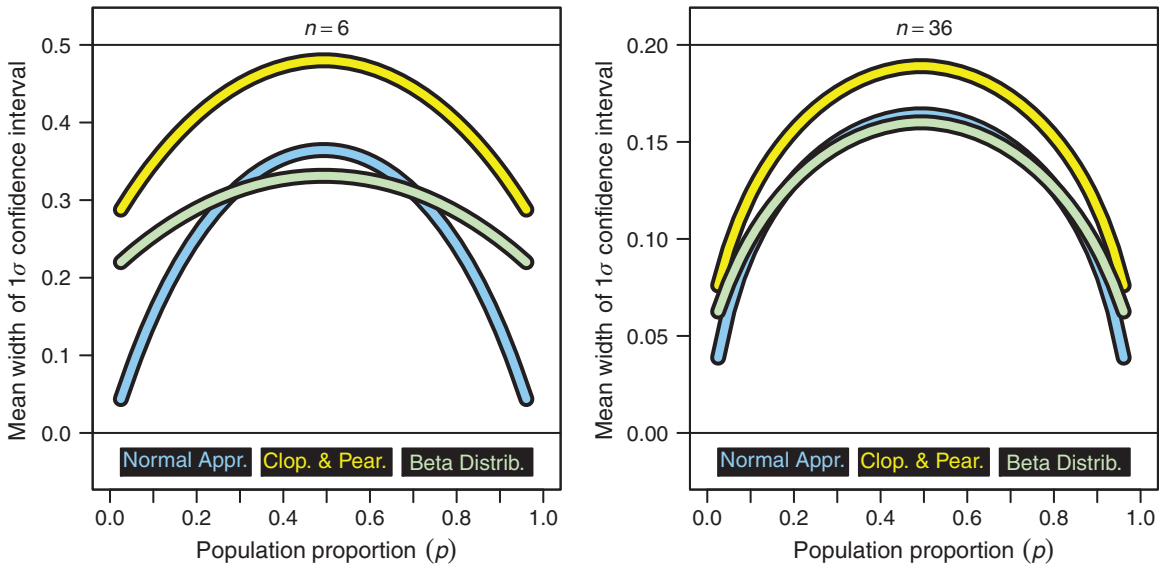

Figure 6 Comparison between the mean widths of binomial CIs generated at $c \approx 0.683$ (1σ) via the beta distribution, the 'normal approximation', and the Clopper & Pearson (1934) approach, respectively, as a function of the underlying population proportion, p, for samples of sizes $n = 6$ (left panel) and $n = 36$ (right panel).

evident in the corresponding *mean* effective coverage for this CI generator, plotted as a function of sample size in the lower panel of Figure 5. Although the Clopper & Pearson (1934) CIs do eventually converge to the nominal level at very large n, in the small to intermediate sample size regime their *mean* effective coverage is consistently far above the desired level. This point is in fact acknowledged in Gehrels (1986), although it appears not to be widely appreciated considering the frequency with which these CIs are treated as a 'gold standard' in astronomical papers.

6 Mean Confidence Interval Widths

To illustrate the influence of the choice of CI generator on the estimated magnitude of the relevant uncertainties (i.e., the error bar size), I compare in Figure 6 the *mean* widths of $c \approx 0.683$ (1σ) CIs estimated via the ('equal-tailed') beta distribution quantile technique, the 'normal approximation', and the Clopper & Pearson (1934) approach as a function of p for samples of sizes $n = 6$ (left panel) and $n = 36$ (right panel). In the small sample size regime (where the 'normal approximation' fails to provide sufficient coverage at $p \lesssim 1/6$ and $p \gtrsim 5/6$; see Figure 4) the mean CI widths are markedly smaller (by as much as $\Delta p \sim -0.15$) than those derived using the beta distribution technique (which generally provides superior coverage at these p values; see Figure 2). (Of course, the beta distribution should not be viewed as a strict benchmark for the ideal CI width, since its coverage is indeed prone to erratic performance at certain p values — the 'oscillation signature' to which *all* non-randomising binomial CI generators are prone; although, as we have argued above, its performance may be considered the best of the three approaches examined in this study.) In the intermediate sample size regime, the mean widths of these these two CI generators are in much better agreement, except at the extremes of $p \lesssim 1/20$ and $p \gtrsim 19/20$ where a marked under-estimation is still evident in the 'normal approximation' CIs. The Clopper & Pearson (1934) CIs, on the other hand, exhibit a much greater mean width than those of the beta distribution or 'normal approximation', regardless of p — reflecting the substantial coverage excess demonstrated for this CI generator in Section 5 (see Figure 5). These examples verify that the choice of CI generator can indeed have a substantial impact on the magnitude of the estimated uncertainties, thereby confirming this choice to be an important practical consideration for effective astronomical data analysis.

7 Conclusions

I have reviewed the performance of three alternative methods for estimating confidence intervals on binomial population proportions; namely, the beta distribution quantile technique, the 'normal approximation', and the Clopper & Pearson (1934) approach (cf. Gehrels 1986). Despite their current popularity in astronomical research, the latter two CI generators are demonstrated to perform poorly under sampling conditions routinely encountered in observational studies, with the 'normal approximation' failing to provide CIs of sufficient width to achieve coverage at the nominal confidence level, and the Clopper & Pearson (1934) approach producing CIs far wider than necessary to achieve the nominal coverage. In contrast, the (Bayesian) beta distribution quantile technique is revealed to be a well-motivated alternative, consistently providing a mean level of coverage close to the nominal level, even for small-to-intermediate sample sizes. Given that the beta distribution generator for binomial CIs may be easily implemented using modern mathematical software packages, I advocate strongly that this technique be adopted in future studies aiming to constrain the true

https://doi.org/10.1071/AS10046 Published online by Cambridge University Press

values of astronomical population proportions (e.g., the galaxy bar fraction, red sequence fraction, or merger fraction).

Acknowledgments

The author would like to thank Matthew Prescott for his assistance in defining the IDL code fragment, Carlos López for supplying the PYTHON code fragment, Roban Kramer for numerous helpful discussions on the role of statistics in astronomy, and the anonymous referee for a thorough reading of the paper and many insightful comments.

References

Agresti, A. & Coull, B. A., 1998, The American Statistician, 52, 2, 119
Baldry, I. K., Balogh, M. L., Bower, R. G., Glazebrook, K., Nicol, R. C., Bamford, S. P. & Budavari, T., 2006, MNRAS, 373, 469
Burgasser, A. J., Kirkpatrick, J. D., Reid, N. I., Brown, M. E., Miskey, C. L. & Gizis, J. E., 2003, ApJ, 586, 512
Brown, L. D., Cai, T. T. & DasGupta, A., 2001, Statistical Science, 16, 2101
Brown, L. D., Cai, T. T. & DasGupta, A., 2002, The Annals of Statistics, 30, 1, 160
Cameron, E. et al., 2010, MNRAS, 409, 1, 346
Clopper, C. J. & Pearson, E. S., 1934, Biometrika, 26, 404
Conselice, C. J., Rajgor, S. & Myers, R., 2008, 386, 909
Cousins, R. D., Hymes, K. E. & Tucker, T., 2009, NIM, 612, 2, 388
De Propris, R., Liske, J., Driver, S. P., Allen, P. D. & Cross, N. J. G., 2005, ApJ, 130, 1516
Elmegreen, D. M., Elmegreen, B. G. & Bellin, A. D., 1990, ApJ, 364, 415
Gehrels, N., 1986, ApJ, 303, 336
Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B., 2003, Bayesian Data Analysis, (New York: Chapman & Hall)
Hester, J. A., 2010, ApJ, 720, 191
Ilbert, O. et al., 2010, ApJ, 709, 644
Kraft, R. P., Burrows, D. N. & Nousek, J. A., 1991, ApJ, 374, 344
Quirin, W. L., 1978, Probability and Statistics (New York: Harper & Row Publishers)
López-Sanjuan, C., Balcells, M., Pérez-González, P. G., Barro, G., Gallego, J. & Zamorano, J., 2010, A&A, 518, 20
Nair, P. B. & Abraham, R. G., 2010, ApJL, 714, 2, L260
Neyman, J., 1935, The Annals of Mathematical Statistics, 6, 111
Rao, M. M. & Swift, R. J., 2006, Mathematics and Its Applications, 582
Ross, T. D., 2003, Computers in Biology and Medicine, 33, 509
Santner, T. J., 1998, Teaching Statistics, 20, 20–23
van den Bergh, S., 2002, AJ, 124, 782
Vollset, S. E., 1993, Statistics in Medicine, 12, 809
Wald, A. & Wolfowitz, J., 1939, The Annals of Mathematical Statistics, 10, 105

A CI Code Fragments & CI Reference Tables

Here I provide simple code fragments demonstrating the implementation of the beta distribution CI generator via standard library routines in R, MATLAB, MATHEMATICA, IDL, and PYTHON. The correct performance of these code fragments in one's preferred mathematical software package may be verified by comparison against the reference tables of binomial CIs presented here in Tables 1 and 2. As in the main body of this paper I denote the nominal confidence level c, the observed success count k, and the sample size n. In the following it is assumed that these variables have already been defined computationally by the user with c a real/double and k and n integers.

In the R statistical package:

```
1. p_lower<-qbeta((1-c)/2,k+1,n-k+1)
2. p_upper<-qbeta(1-(1-c)/2,k+1,n-k+1)
```

In MATLAB:

```
1. p_lower=betaincinv((1-c)/2,k+1,
   n-k+1)
2. p_upper=betaincinv(1-(1-c)/2,k+1,
   n-k+1)
```

In MATHEMATICA:

```
1. plower=Quantile[BetaDistribution
   [k+1,n-k+1],(1-c)/2]
2. pupper=Quantile[BetaDistribution
   [k+1,n-k+1],1-(1-c)/2]
```

In IDL (if an 'IDL Analyst' license is available):

```
1. p_lower=IMSL_BETACDF((1-c)/2,k+1,
   n-k+1,/INVERSE)
2. p_upper=IMSL_BETACDF(1-(1-c)/2,
   k+1,n-k+1,/INVERSE) otherwise, iteratively:
3. z=FINDGEN(10000)*0.0001
4. Beta=IBETA(k+1,n-k+1,z)
5. il=VALUE_LOCATE(Beta,(1-c)/2)
6. iu=VALUE_LOCATE(Beta,1-(1-c)/2)
7. p_lower=z[il]
8. p_upper=z[iu]
```

In PYTHON:

```
1. import scipy.stats.distributions as
   dist
2. p_lower=dist.beta.ppf((1-c)/2.,
   k+1,n-k+1)
3. p_upper=dist.beta.ppf(1-(1-c)/2.,
   k+1,n-k+1)
```

https://doi.org/10.1071/AS10046 Published online by Cambridge University Press

Table 1. Confidence interval estimates at $c \approx 0.683$ (1σ) on binomial population proportions from quantiles of the beta distribution for all possible observed success counts for sample sizes up to 20

n	$k=0$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.602	0.917	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
	0.083	0.398																			
2	0.459	0.748	0.944	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
	0.056	0.252	0.541																		
3	0.369	0.618	0.815	0.958	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
	0.042	0.185	0.382	0.631																	
4	0.308	0.524	0.703	0.853	0.966	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
	0.034	0.147	0.297	0.476	0.692																
5	0.264	0.454	0.615	0.757	0.879	0.972	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
	0.028	0.121	0.243	0.385	0.546	0.736															
6	0.231	0.400	0.546	0.676	0.794	0.896	0.976	…	…	…	…	…	…	…	…	…	…	…	…	…	…
	0.024	0.104	0.206	0.324	0.454	0.600	0.769														
7	0.206	0.357	0.490	0.610	0.720	0.821	0.910	0.979	…	…	…	…	…	…	…	…	…	…	…	…	…
	0.021	0.090	0.179	0.280	0.390	0.510	0.643	0.794													
8	0.185	0.323	0.444	0.555	0.658	0.754	0.842	0.920	0.981	…	…	…	…	…	…	…	…	…	…	…	…
	0.019	0.080	0.158	0.246	0.342	0.445	0.556	0.677	0.815												
9	0.168	0.294	0.405	0.508	0.605	0.695	0.780	0.858	0.928	0.983	…	…	…	…	…	…	…	…	…	…	…
	0.017	0.072	0.142	0.220	0.305	0.395	0.492	0.595	0.706	0.832											
10	0.154	0.270	0.373	0.469	0.559	0.644	0.725	0.801	0.872	0.935	0.984	…	…	…	…	…	…	…	…	…	…
	0.016	0.065	0.128	0.199	0.275	0.356	0.441	0.531	0.627	0.730	0.846										
11	0.142	0.250	0.346	0.435	0.519	0.600	0.676	0.750	0.819	0.883	0.940	0.986	…	…	…	…	…	…	…	…	…
	0.014	0.060	0.117	0.181	0.250	0.324	0.400	0.481	0.565	0.654	0.750	0.858									
12	0.132	0.232	0.322	0.405	0.485	0.561	0.634	0.703	0.770	0.833	0.892	0.945	0.987	…	…	…	…	…	…	…	…
	0.013	0.055	0.108	0.167	0.230	0.297	0.366	0.439	0.515	0.595	0.678	0.768	0.868								
13	0.123	0.217	0.301	0.380	0.455	0.526	0.595	0.662	0.726	0.787	0.846	0.900	0.949	0.988	…	…	…	…	…	…	…
	0.012	0.051	0.100	0.154	0.213	0.274	0.338	0.405	0.474	0.545	0.620	0.699	0.783	0.877							
14	0.116	0.204	0.283	0.357	0.428	0.496	0.561	0.625	0.686	0.745	0.802	0.856	0.907	0.952	0.989	…	…	…	…	…	…
	0.011	0.048	0.093	0.144	0.198	0.255	0.314	0.375	0.439	0.504	0.572	0.643	0.717	0.796	0.884						
15	0.109	0.192	0.267	0.337	0.404	0.469	0.531	0.592	0.650	0.707	0.762	0.815	0.866	0.913	0.955	0.989	…	…	…	…	…
	0.011	0.045	0.087	0.134	0.185	0.238	0.293	0.350	0.408	0.469	0.531	0.596	0.663	0.733	0.808	0.891					
16	0.103	0.181	0.252	0.319	0.383	0.444	0.504	0.562	0.618	0.673	0.726	0.777	0.826	0.874	0.918	0.958	0.990	…	…	…	…
	0.010	0.042	0.082	0.126	0.174	0.223	0.274	0.327	0.382	0.438	0.496	0.556	0.617	0.681	0.748	0.819	0.897				
17	0.097	0.172	0.239	0.303	0.363	0.422	0.479	0.534	0.588	0.641	0.692	0.742	0.790	0.836	0.881	0.923	0.960	0.990	…	…	…
	0.010	0.040	0.077	0.119	0.164	0.210	0.258	0.308	0.359	0.412	0.466	0.521	0.578	0.637	0.697	0.761	0.828	0.903			
18	0.092	0.163	0.228	0.288	0.346	0.402	0.456	0.509	0.561	0.612	0.661	0.709	0.756	0.802	0.845	0.887	0.927	0.962	0.991	…	…
	0.009	0.038	0.073	0.113	0.155	0.198	0.244	0.291	0.339	0.388	0.439	0.491	0.544	0.598	0.654	0.712	0.772	0.837	0.908		
19	0.088	0.156	0.217	0.275	0.330	0.384	0.436	0.487	0.537	0.585	0.633	0.679	0.725	0.769	0.812	0.853	0.893	0.930	0.964	0.991	…
	0.009	0.036	0.070	0.107	0.147	0.188	0.231	0.275	0.321	0.367	0.415	0.463	0.513	0.564	0.616	0.670	0.725	0.783	0.844	0.912	
20	0.084	0.149	0.207	0.263	0.316	0.367	0.417	0.466	0.514	0.561	0.607	0.652	0.696	0.739	0.780	0.821	0.861	0.898	0.934	0.966	0.992
	0.008	0.034	0.066	0.102	0.139	0.179	0.220	0.261	0.304	0.348	0.393	0.439	0.486	0.534	0.583	0.633	0.684	0.737	0.793	0.851	0.916

Table 2. Confidence interval estimates at $c \approx 0.997$ (3σ) on binomial population proportions from quantiles of the beta distribution for all possible observed success counts for sample sizes up to 20

n	$k=0$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.963	0.999	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
	0.001	0.037																			
2	0.889	0.979	1.000	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
	0.000	0.021	0.111																		
3	0.808	0.929	0.985	1.000	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
	0.000	0.015	0.071	0.192																	
4	0.733	0.868	0.947	0.988	1.000	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
	0.000	0.012	0.053	0.132	0.267																
5	0.668	0.807	0.898	0.958	0.990	1.000	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
	0.000	0.010	0.042	0.102	0.193	0.332															
6	0.611	0.750	0.847	0.917	0.965	0.992	1.000	…	…	…	…	…	…	…	…	…	…	…	…	…	…
	0.000	0.008	0.035	0.083	0.153	0.250	0.389														
7	0.562	0.698	0.797	0.872	0.930	0.970	0.993	1.000	…	…	…	…	…	…	…	…	…	…	…	…	…
	0.000	0.007	0.030	0.070	0.128	0.203	0.302	0.438													
8	0.520	0.652	0.750	0.828	0.891	0.939	0.974	0.994	1.000	…	…	…	…	…	…	…	…	…	…	…	…
	0.000	0.006	0.026	0.061	0.109	0.172	0.250	0.348	0.480												
9	0.484	0.610	0.707	0.785	0.851	0.904	0.946	0.977	0.994	1.000	…	…	…	…	…	…	…	…	…	…	…
	0.000	0.006	0.023	0.054	0.096	0.149	0.215	0.293	0.390	0.516											
10	0.452	0.573	0.667	0.745	0.812	0.868	0.915	0.952	0.979	0.995	1.000	…	…	…	…	…	…	…	…	…	…
	0.000	0.005	0.021	0.048	0.085	0.132	0.188	0.255	0.333	0.427	0.548										
11	0.423	0.540	0.632	0.708	0.775	0.832	0.882	0.923	0.956	0.981	0.995	1.000	…	…	…	…	…	…	…	…	…
	0.000	0.005	0.019	0.044	0.077	0.118	0.168	0.225	0.292	0.368	0.460	0.577									
12	0.398	0.510	0.599	0.674	0.740	0.798	0.848	0.893	0.930	0.960	0.982	0.996	1.000	…	…	…	…	…	…	…	…
	0.000	0.004	0.018	0.040	0.070	0.107	0.152	0.202	0.260	0.326	0.401	0.490	0.602								
13	0.376	0.484	0.569	0.643	0.707	0.765	0.816	0.862	0.902	0.936	0.963	0.984	0.996	1.000	…	…	…	…	…	…	…
	0.000	0.004	0.016	0.037	0.064	0.098	0.138	0.184	0.235	0.293	0.357	0.431	0.516	0.624							
14	0.356	0.459	0.542	0.614	0.677	0.734	0.785	0.832	0.873	0.909	0.941	0.966	0.985	0.996	1.000	…	…	…	…	…	…
	0.000	0.004	0.015	0.034	0.059	0.091	0.127	0.168	0.215	0.266	0.323	0.386	0.458	0.541	0.644						
15	0.338	0.438	0.517	0.587	0.649	0.705	0.756	0.802	0.845	0.882	0.916	0.945	0.968	0.986	0.997	1.000	…	…	…	…	…
	0.000	0.003	0.014	0.032	0.055	0.084	0.118	0.155	0.198	0.244	0.295	0.351	0.413	0.483	0.562	0.662					
16	0.322	0.417	0.495	0.562	0.623	0.678	0.728	0.774	0.817	0.856	0.891	0.922	0.948	0.970	0.987	0.997	1.000	…	…	…	…
	0.000	0.003	0.013	0.030	0.052	0.078	0.109	0.144	0.183	0.226	0.272	0.322	0.377	0.438	0.505	0.583	0.678				
17	0.307	0.399	0.474	0.539	0.598	0.652	0.702	0.748	0.790	0.829	0.865	0.898	0.926	0.952	0.972	0.988	0.997	1.000	…	…	…
	0.000	0.003	0.012	0.028	0.048	0.074	0.102	0.135	0.171	0.210	0.252	0.298	0.348	0.402	0.461	0.526	0.601	0.693			
18	0.294	0.382	0.455	0.518	0.575	0.628	0.677	0.722	0.765	0.804	0.840	0.874	0.904	0.931	0.954	0.974	0.988	0.997	1.000	…	…
	0.000	0.003	0.012	0.026	0.046	0.069	0.096	0.126	0.160	0.196	0.235	0.278	0.323	0.372	0.425	0.482	0.545	0.618	0.706		
19	0.281	0.367	0.437	0.498	0.554	0.606	0.654	0.698	0.740	0.779	0.816	0.850	0.881	0.909	0.935	0.957	0.975	0.989	0.997	1.000	…
	0.000	0.003	0.011	0.025	0.043	0.065	0.091	0.119	0.150	0.184	0.221	0.260	0.302	0.346	0.394	0.446	0.502	0.563	0.633	0.719	
20	0.270	0.353	0.420	0.480	0.535	0.585	0.632	0.676	0.717	0.756	0.792	0.826	0.858	0.887	0.914	0.938	0.959	0.976	0.989	0.997	1.000
	0.000	0.003	0.011	0.024	0.041	0.062	0.086	0.113	0.142	0.174	0.208	0.244	0.283	0.324	0.368	0.415	0.465	0.520	0.580	0.647	0.730