The Statistician (1990) 39, pp. 391–397

Exact confidence limits for binomial proportions—Pearson and Hartley revisited

D. J. BRENNER & H. QUAN

Center for Radiological Research and Department of Statistics, Columbia University, 630 W 168th Street, New York, NY 10032, USA

Abstract. It has long been known that there is a problem with the results of 'conventional' (as epitomized in the *Biometrika Tables for Statisticians*) techniques for calculating confidence limits for parameters of discrete distributions. Specifically, the calculated limits at a given confidence level are always much too wide, i.e. overly conservative. For large sample sizes ($n \geqslant 100$) this is not important, but for small samples, the conventional techniques can be very conservative. In this note, exact confidence limits for the parameter p, as calculated in binomial sampling, are presented. A Bayesian technique is used, and the results are presented for the situation where no prior information is assumed, corresponding to the 'conventional' scenario for confidence-limit estimation. Our results are compared quantitatively with those in the Biometrika Tables by use of Monte-Carlo simulation. The results show, as expected, that for small sample sizes, the Biometrika Tables yield confidence intervals that are considerably too wide, and that our approach does indeed produce exact confidence limits. Extensive graphs of results are presented.

1 Introduction

A common situation is for observations to be binomially distributed with a probability density given by:

$$f(i|n,p)=\binom{n}{i}p^i(1-p)^{n-i} \qquad (i=0,1,\ldots,n). \tag{1.1}$$

Here n is the sample size, i the observed frequency and p the binomial proportion. It is often of importance to compute lower and upper $(1-2\alpha)$ confidence limits for p, $p_A(c|n,\alpha)$ and $p_B(c|n,\alpha)$, on the basis of an observed value $i=c$.

It has been known for over half a century that, for small values of c or $n-c$, there is a problem with such confidence-limit estimation, because of the discrete nature of f; both Clopper & Pearson (1934) and Fisher (1935) made this observation. Indeed in the standard *Biometrika Tables for Statisticians* (Pearson & Hartley, 1970), the comment is made that "the probability that the statement

$$p_A(c|n,\alpha) \leqslant p \leqslant p_B(c|n,\alpha) \tag{1.2}$$

is correct is likely to be in considerable excess of the lower bound, 1–2α." In this note we present results for p_A and p_B which correspond exactly to $1-2\alpha$ confidence limits.

2 Methods for calculating p_A and p_B

The 'standard' technique for estimating p_A and p_B (e.g. Pearson & Hartley, 1970) is to solve

$$\sum_{i=c}^{n} f(i|n,p_A)=\alpha \tag{2.1}$$

$$\sum_{i=0}^{c} f(i|n,p_B)=\alpha \tag{2.2}$$

for p_A and p_B. The continuity problem arises because, given p, the discreet nature of the binomial distribution does not in general allow integer solutions, c_1 and c_2, to the equations

$$\sum_{i=c_1}^{n} f(i|n,p)=\alpha \tag{2.3}$$

$$\sum_{i=0}^{c_2} f(i|n,p)=\alpha. \tag{2.4}$$

In practice, the technique is to make c_1 as small as possible (and c_2 as large as possible) subject to the corresponding sums in (2.3) and (2.4) not exceeding α. This yields probabilities of at most α that $c \geqslant c_1$ and $c \leqslant c_2$ or, equivalently, a probability of at least $1-2\alpha$ that (1.2) is correct. Thus the technique leads to overly conservative (i.e. too wide) estimates of the interval $[p_A, p_B]$.

Exact confidence limits for the binomial proportion can be calculated using a Bayesian approach (Lindley, 1965; Arnett, 1976; Brenner & Quan, 1990); Bayes' theorem gives the density distribution, g, for p given a measurement c, a sample size n and a density distribution, h, for any prior information that might be available:

$$g(p|n,c,h)=\frac{h(p)f(c|n,p)}{\int h(p')f(c|n,p')\,\mathrm{d}p'}. \tag{2.5}$$

A convenient form for the prior distribution, $h(p)$, is the power function

$$h(p)=(1+\beta)p^{\beta}, \tag{2.6}$$

which is a special case of the more general beta distribution. Note that $\beta=0$ corresponds to the situation where no prior information is assumed. The upper and lower confidence limits for p are then the limits of the integral

$$\int_{p_A}^{p_B} g(p|n,c,h)\,\mathrm{d}p=1-2\alpha. \tag{2.7}$$

The added, reasonable, constraint that the length of the interval $[p_A, p_B]$ should be minimized, leads (Lindley, 1965) to a second equation

$$g(p_A|n,c,h)=g(p_B|n,c,h). \tag{2.8}$$

Equations (2.7) and (2.8) can be numerically solved for a given set of $[\alpha, \beta, c, n]$ to yield $[p_A, p_B]$. We have performed such calculations for the case where $\beta=0$ (implying no prior information about p), in order to facilitate comparison with the results (hereafter referred to as P&H) of solving (2.1) and (2.2)—which also involve no prior information about p.

3 Results

Figure 1 shows a comparison of p_A and p_B (99% confidence limits) for various sample sizes, n, and measurements, c. p_A and p_B are calculated either using the P&H technique or by using (2.7) and (2.8) with $\beta=0$. It is clear that for small values of n the results are distinctly different.

In order to quantify the significance of the differences between our results and those obtained using the P&H technique, a Monte-Carlo simulation was used. The technique was

(A) Choose a random number, P, with density distribution $h(P)$, between 0 and 1. For $\beta=0$, we choose P uniformly between 0 and 1.

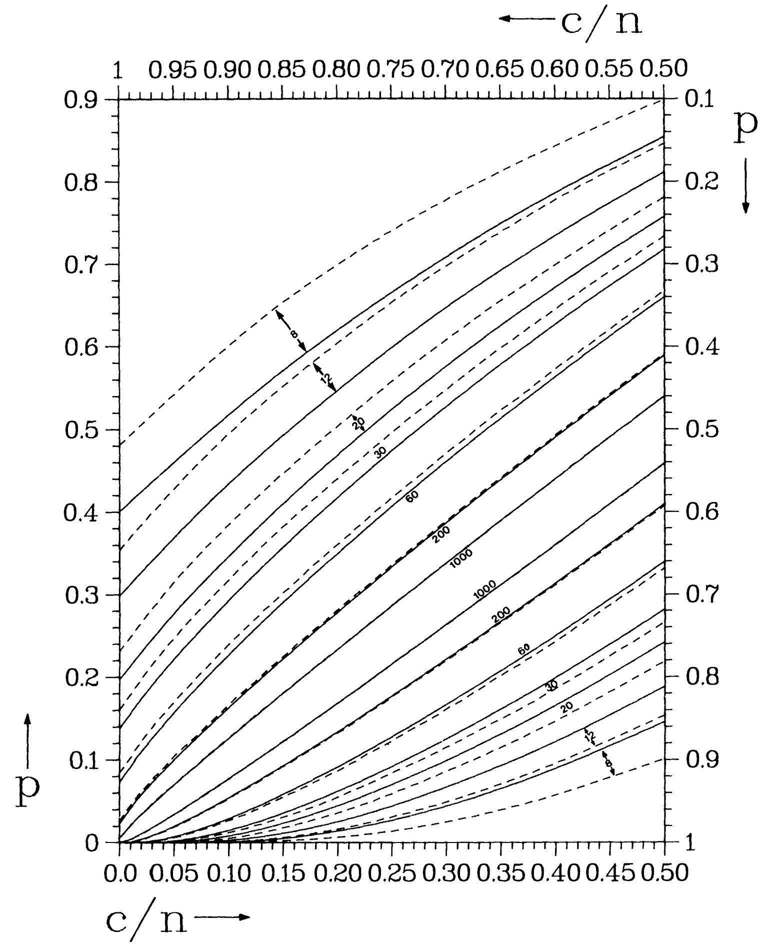

Fig. 1. The 99% confidence limits for the binomial proportion, p; the upper and lower sets of curves are for the upper and lower limits, p_B and p_A, respectively. The labels on the curves refer to the sample size, n, and the abscissa is the ratio of the measurement, c, to n. The dashed curves are the approximate, conservative, limits as calculated with (2.1) and (2.2). The full curves are exact limits calculated with (2.7) and (2.8). The full curves are reproduced in Fig. 4 for greater clarity.

(B) Calculate the probability, f, (see 1.1) of obtaining a measurement c out of a sample size n and binomial proportion P.
(C) With probability f, add P to a frequency distribution $F(P')$; this is done by comparing f with a second random number, Q, uniformly distributed from 0 to 1. If $Q < f$, then P is added to $F(P')$ with unit probability.
(D) Repeat (A)–(C) 25 000 times.

The results of the simulation confirmed our expectations. As an example, for a sample size $n = 8$, and an observed count $c = 4$, the 99% confidence interval for p as calculated using P&H is [0·101, 0·899], whilst using (2.7) and (2.8) it is [0·146, 0·854]; using the Monte-Carlo simulation, P fell within these limits 99·8% and 99·0% of the time, respectively, confirming that P&H gives conservative confidence limits, while the current approach yields exact limits. (Several repetitions of this procedure with different random number starting seeds indicated that the uncertainty in these proportions due to the Monte-Carlo technique itself was considerably less than 0·1%.)

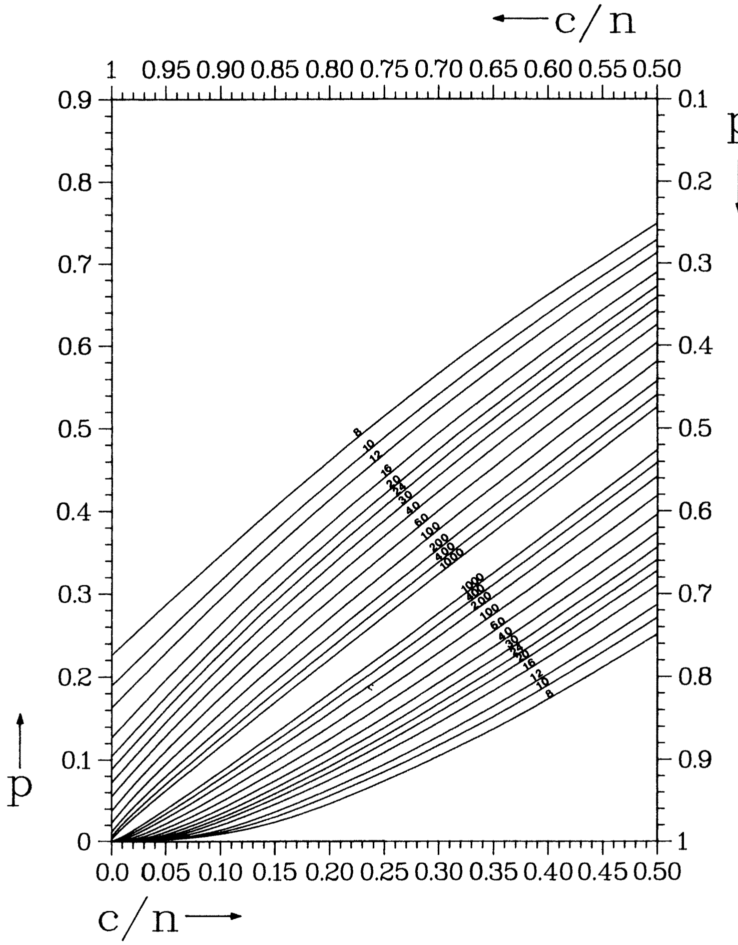

Fig. 2. Exact 90% confidence limits for the binomial proportion, *p*. The calculations were performed using (2.7) and (2.8).

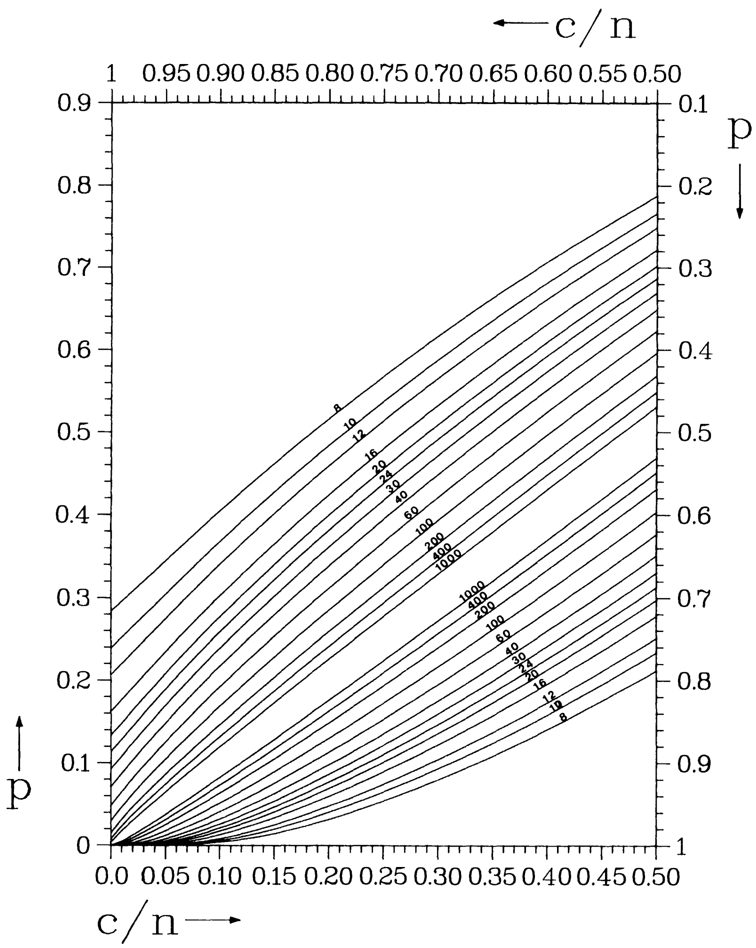

Fig. 3. As Fig. 2 for 95% confidence level.

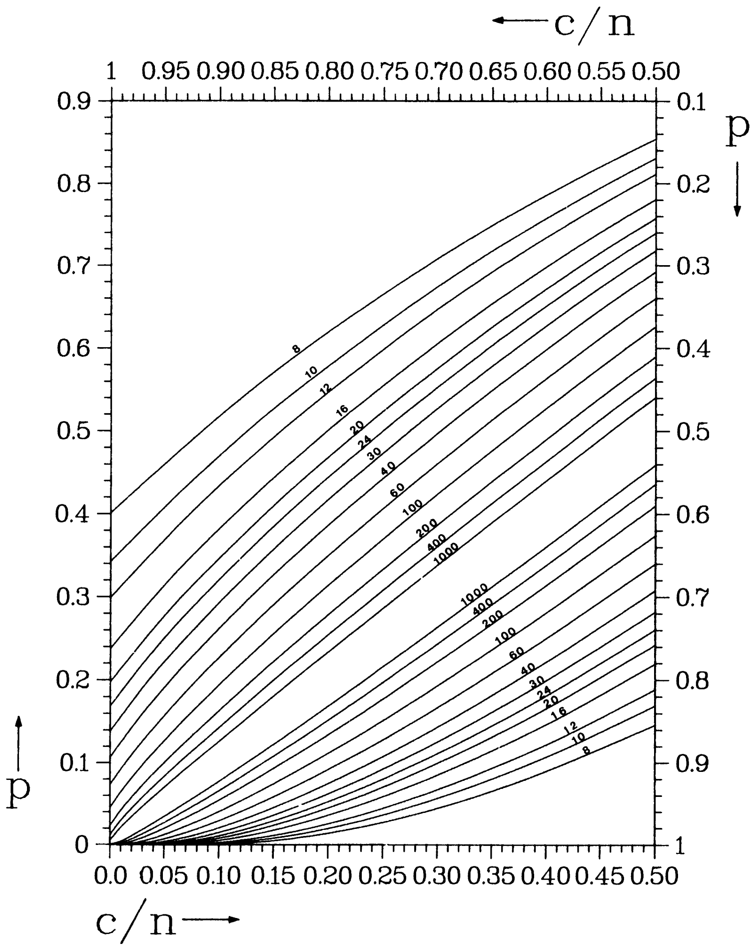

Fig. 4. As Fig. 2 for 99% confidence level.

As expected, the difference between these two approaches decreases as n increases. For example for $n=60$ and $c=30$, the calculated confidence limits are [0·332, 0·668] (P&H) and [0·340, 0·660] (this approach). Using the Monte-Carlo simulation, P fell within these limits 99·3% and 99·0% of the time, respectively.

Finally, for reference, we show in Figs 2–4 graphs analogous to those in Pearson & Hartley (1970) for 90%, 95% and 99% confidence limits as calculated using the current approach.

Acknowledgements

The useful comments made by the reviewer of this paper are gratefully acknowledged. This work was supported by Grants CA-12536 from the National Cancer Institute and DE-FG02-88ER60631 from the US Department of Energy.

References

ARNETT, L. M. (1976) Confidence limits for parameters of Poisson and binomial distributions. Du Pont Report DP-1416.

BRENNER, D. J. & QUAN, H. (1990) Confidence limits for low induced frequencies of oncogenic transformation in the presence of a background, *International Journal of Radiation Biology*, 57, pp. 1031–1045.

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

FISHER, R. A. (1935) The logic of inductive inference, *Journal of the Royal Statistical Society*, 98, pp. 39–82.

LINDLEY, D. V. (1965) *Introduction to Probability and Statistics from a Bayesian Viewpoint* (Cambridge, Cambridge University Press).

PEARSON, E. S. & HARTLEY, H. O. (1970) *Biometrika Tables for Statisticians*, Vol. 1 (Cambridge, Cambridge University Press).