

Confidence Intervals for Binomial Distributions Using SAS Graphics and PROBBNML

Alan Davis, Ph.D.

Department of Family Medicine, University of Oklahoma, Oklahoma
City

Abstract

An iterative program is shown which determines confidence intervals for the estimate of a binomial probability parameter PI. Comparisons are made with the normal approximation of the confidence interval. For $n=20$, $\alpha=.05$, graphs show the characteristics of these confidence intervals. A second program is provided for determining the expected degree of confidence for specific values of PI. These programs can be run in a SAS (R) data step and can be easily edited to any n or α .

Introduction

The binomial probability coefficient PI is the probability that an event will happen in each of n independent trials. A confidence interval estimate of the binomial probability coefficient from a sample of size n is an interval (PL,PU) within the interval (0,1) which contains the true parameter with some degree of certainty α . Thus a 95% confidence interval would contain the true parameter with probability:

$$1 - \alpha = .95.$$

The confidence interval is constructed by ruling out those parameters with low probability of producing an observed sample of data. Consider a sample containing m occurrences of the random event, and the estimate of PI is:

$$\text{PHAT} = m/n.$$

The normal approximation (Braverman and Stewart, 1973) is often used to construct confidence intervals for the binomial, which would be the interval $\text{PHAT} \pm (1.96 * \text{SQRT}(\text{PQ}/(N-1)))$. For values of p other than .5 or for $np < 5$ the poisson approximation to the binomial distribution is recommended, but this method is limited to fractions having n in the denominator.

Using PROBBNML(p,n,m), where p is an estimate of the confidence limit, n is the sample size, and m is the number of occurrences, an iterative search technique finds that p which will be arbitrarily close in probability to the desired confidence limit. E.g., the lower confidence bound PL of a two-sided confidence interval estimate would be that for which

$$(1 - \text{PROBBNML}(\text{PL},n,m) - \alpha/2) \text{ is small.}$$

The upper confidence bound PU is that for which

($\text{PROBBNML}(\text{PU}, n, m) - \alpha/2$) is small.

The Program which produces confidence interval estimates uses the variable CRIT as its criterion for smallness. As an initial estimate, the normal approximation confidence limits are used. This program is as follows:

```
data confint; N=20; alpha=.05; zalpha=1.96; a2=alpha/2;
```

```
do m=0 to n;
```

```
  phat=m/n;
```

```
  p0=phat;
```

```
  n1=phat-zalpha*sqrt(phat*(1-phat)/(n-1));
```

```
  if n1 < 0 then n1=0;
```

```
  p1=n1;
```

```
  a0= 1 - probbnml(p0,n,m-1);
```

```
lower:
```

```
  a1=1-probbnml(p1,n,m-1);
```

```
  if abs(a0-a1) < .00001 then go to upper;
```

```
  p2=p1+(p1-p0)*(a2-a1)/(a1-a0);
```

```
  if p2 > 1 then p2=1; if p2 < 0 then p2=0;
```

```
  a0=a1; p0=p1; p1=p2;
```

```
  go to lower;
```

```
upper:      p1=p2;
```

```
  p0=phat;
```

```
  nu=phat+zalpha*sqrt(phat*(1-phat)/(n-1));
```

```
  if nu > 1 then nu=1;
```

```
  p1=nu;
```

```
  a0=probbnml(p0,n,m);
```

```
upper2:
```

```
  a1=probbnml(p1,n,m);
```

```
  if abs(a0-a1) < .00001 then go to next;
```

```
  p2=p1+(p1-p0)*(a2-a1)/(a1-a0);
```

```
  if p2 > 1 then p2=1; if p2 < 0 then p2=0;
```

```
  a0=a1; p0=p1; p1=p2;
```

```
  go to upper2;
```

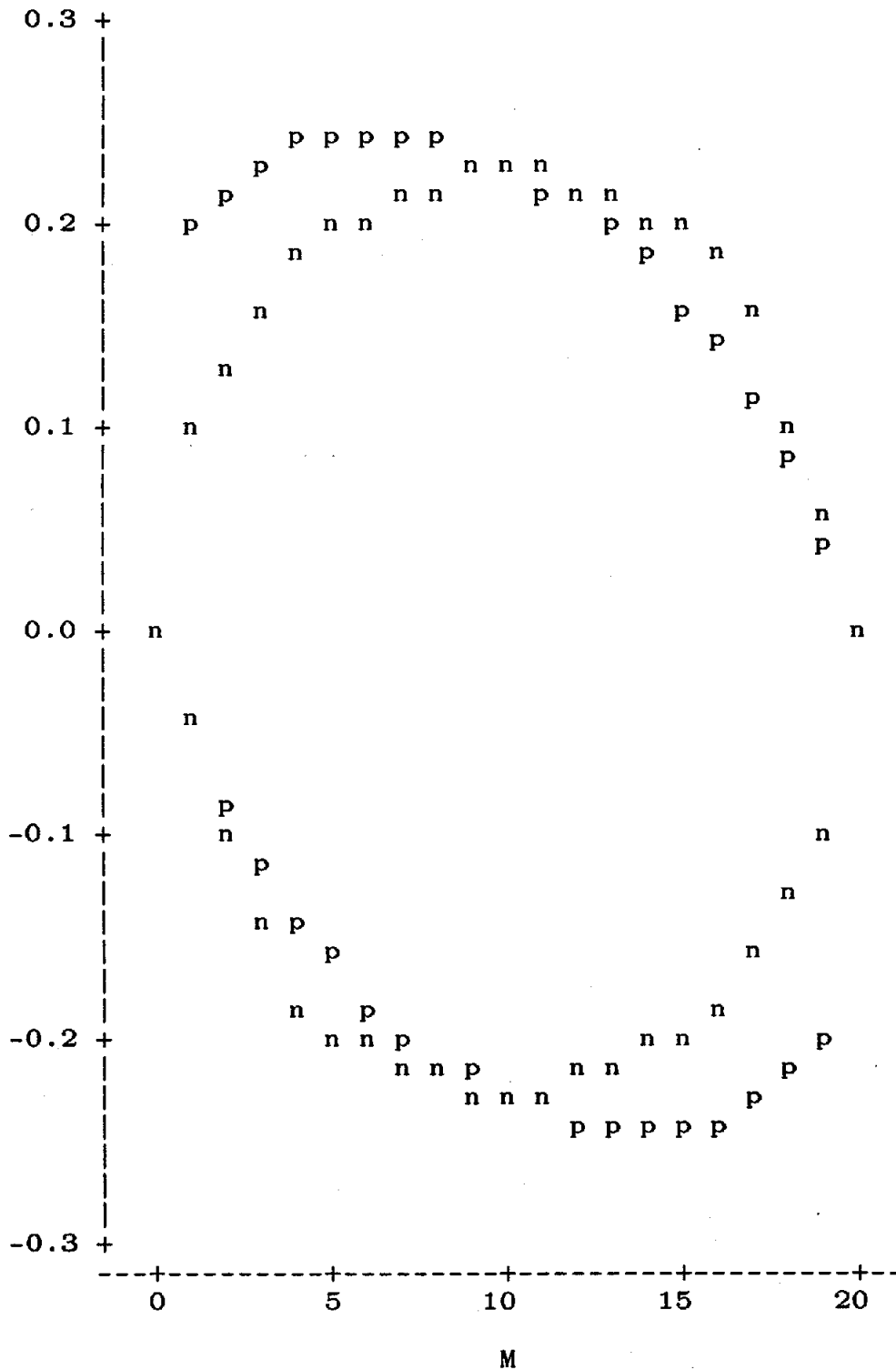
```
next:      pu=p2;
```

```
  output;
```

```
  end;
```

```
run;
```

FIGURE 1. Plot of interval to upper limit (positive) and lower limit (negative) for PROBBNML estimate (p) and Normal distribution estimate (n) for upper and lower confidence limits, versus m, an observed number of successes in $n = 20$ trials.



NOTE: 11 obs hidden.

Figure 1 indicates some large differences between the normal approximation and the iterative technique using PROBBNML. The largest difference is .1 at the extreme of $m=1$. The general appearance of the curves is due to the difference in deviation for a binomial versus the normal approximation. The upper limit is larger for the binomial where $M < 10$, because the variation of the binomial increases with larger values until it reaches .5. After that point, larger values of P have smaller variance, so the binomial upper limit becomes less than the normal.

For the lower limit of the confidence intervals, the binomial uses smaller parameters having smaller variance than the variance used in the normal approximation. This reduces the width of the confidence interval of the binomial relative to the normal approximation.

Comparison of Confidence Intervals for Specific PI values.

Given a specific value of PI, the sum of probability for $m=0, \dots, n$ for those intervals containing the true parameter yields the expected degree of confidence. A program, using PROBBNML, to compare the normal with the binomial techniques follows:

```
data ;
do pi = .1, .25, .5, .75, .9; keep n pi nconf pconf;
  nconf=0; pconf=0; totprob=0;
  label nconf='normal approximation' pconf='binomial estimate'
  ;
  *determine for several true binom pi what true conf interval
  is;
  plast=0;
  do recno= 1 to nob;
    SET confint POINT = RECNO NOBS=NOBS;
    pm = probbnml(pi,n,m);
    prob= pm-plast;
    plast=pm;
    if (pi-nl)*(pi-nu)<=0 then nconf+prob;
    if (pi-pl)*(pi-pu)<=0 then pconf+prob;
    totprob+prob;
  end; *of input loop;
  output;
end; *of loop over pi;
stop;
proc print label;
run;
```

Comparisons for specific values of PI are shown in the following table:

OBS	PI	normal approximation	binomial estimate	N
1	0.10	0.87604	0.86717	20
2	0.25	0.89488	0.96182	20
3	0.50	0.95861	0.95861	20
4	0.75	0.89488	0.96182	20
5	0.90	0.87604	0.86717	20

Curiously, the points $PI = .1$ and $.9$ are closer to the nominal 0.95 confidence interval, despite the fact that the binomial confidence interval is larger. As an explanation, consider figure 2. Each of the confidence intervals is displayed horizontally, and the test values for PI are highlighted with vertical lines. The increase of confidence at $.1$ and $.9$ are due to the intersection with the normal confidence interval and a near miss of the binomial for the same value of m .

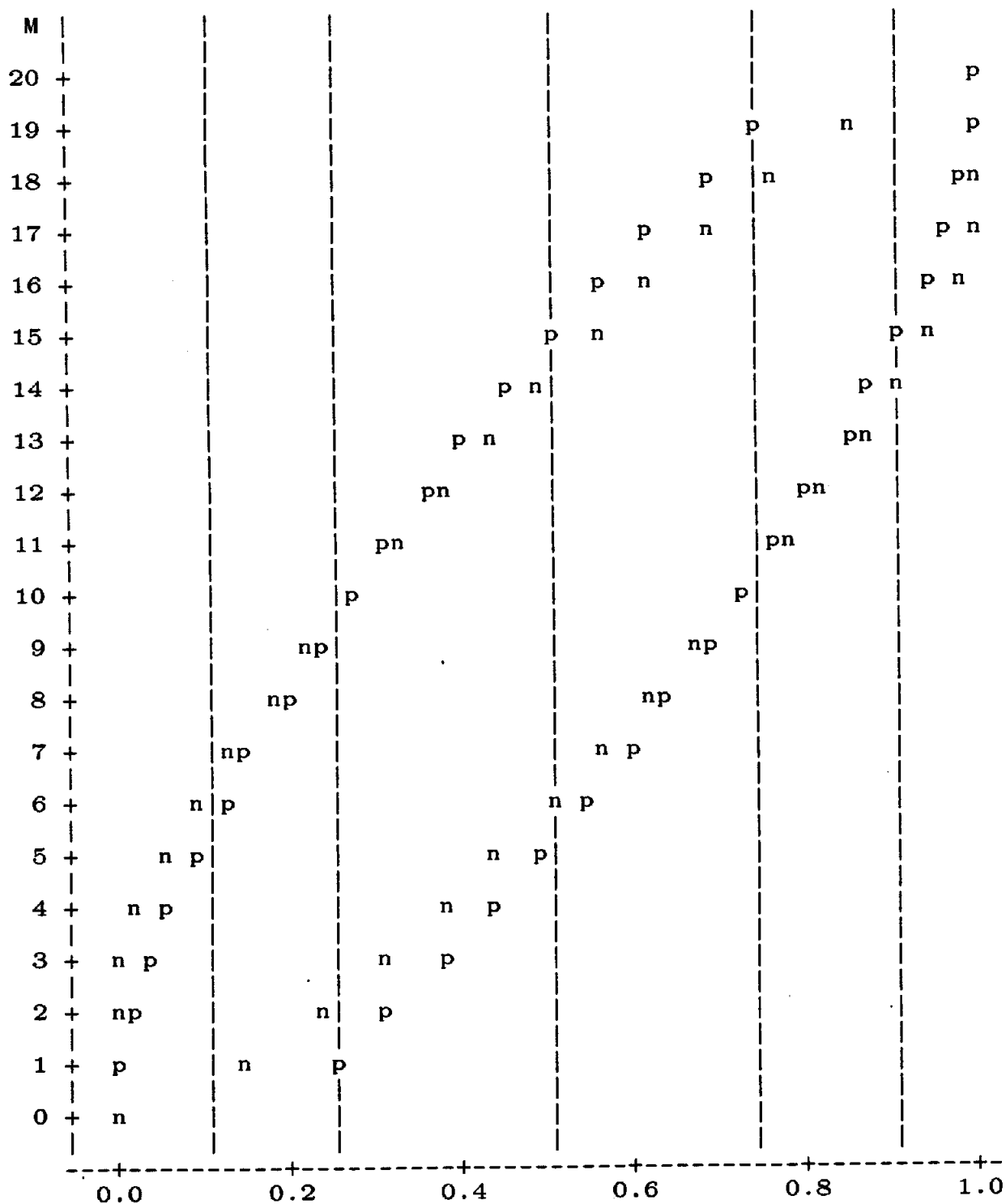
Figure 3 shows the symmetrical nature of the confidence intervals.

Conclusion

These programs can be used to calculate and test confidence intervals, and help the student and statistical consultant visualize binomial confidence intervals. Only the Basic SAS (R) software is required.

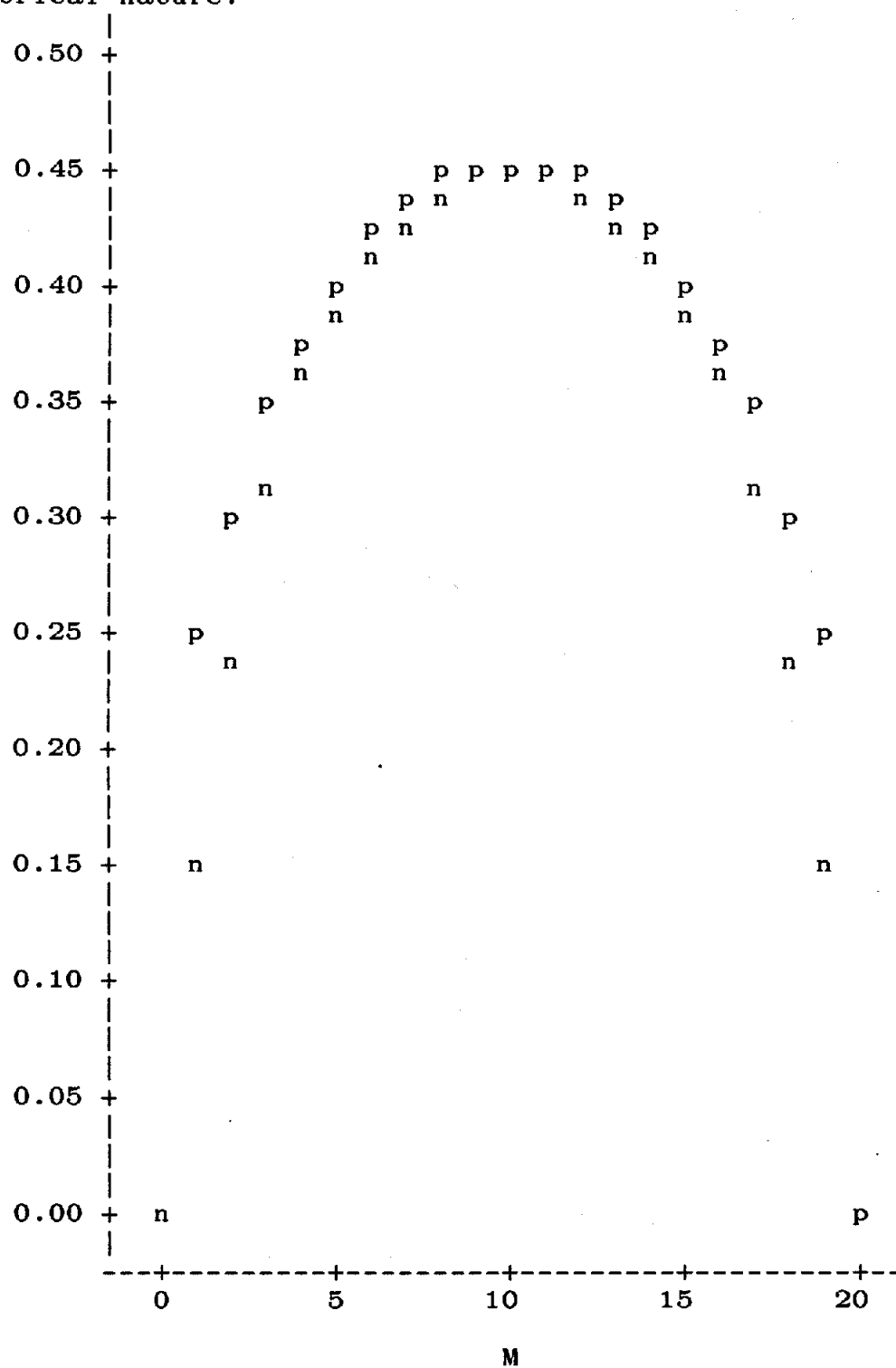
Braverman, J.D., and Stewart, W.C., Statistics for Business and Economics Copyright 1973 by Ronald Press Co.
8:28 Thursday, August 29, 1991

Figure 2. Intersection of Confidence intervals (p = binomial
 n =normal approximation) with specific values of PI.



NOTE: 10 obs hidden.

Figure 3. Length of confidence intervals showing symmetrical nature.



NOTE: 1 obs had missing values. 4 obs hidden.