Coverage versus Confidence
A Tutorial

Peyton Cook

This article is intended to help students understand the concept of a coverage probability involving confidence intervals. Mathematica is used as a language for describing an algorithm to compute the coverage probability for a simple confidence interval based on the binomial distribution. Then, higher-level functions are used to compute probabilities of expressions in order to obtain coverage probabilities. Several examples are presented: two confidence intervals for a population proportion based on the binomial distribution, an asymptotic confidence interval for the mean of the Poisson distribution, and an asymptotic confidence interval for a population proportion based on the negative binomial distribution.

■ 1. Introduction

Introductory courses in mathematical statistics present the rudimentary concepts behind confidence intervals. The creation of confidence intervals often involves the use of maximum likelihood estimation and the central limit theorem along with estimated standard errors. This is described in Casella and Berger [1, p. 497]. Consequently, the level of confidence is often only approximate. This is particularly the case when continuous probability models are used to approximate discrete probabilities. The probability that the interval surrounds the unknown parameter depends on the value of the unknown parameter. Such a probability is called a *coverage probability*. Confidence is defined as the infimum of the coverage probabilities. The following definitions can be found in Casella and Berger [1, p. 418].

Definition (Coverage and Confidence)

Let $X = (X_1, X_2, \ldots, X_n)$ where the X_i are all independent from a distribution with probability density (or discrete mass) function given by $f(x; \theta), x \in \chi, \theta \in \Theta$. The support of each X_i is χ and the parameter space is Θ. Let $L(X)$ and $U(X)$ be the lower and upper limits of a confidence interval. Then the coverage probability of the interval evaluated at θ is coverage$(\theta) = P(L(X) < \theta < U(X); \theta)$. The level of confidence is $\inf_{\theta \in \Theta}$ coverage(θ).

Students are often confused about how to compute coverage probabilities. This tutorial is intended to help students understand them. We give a detailed explanation of calculating one particular coverage probability. This also allows one to perform the calculations with a minimum of distraction involving programming. We then compute coverage probabilities using higher-level functions `Probability` and `NProbability` that allow specifying a function of a random variable along with its distribution. In both cases these functions allow one to focus on the higher-level ideas rather than low-level nuts and bolts of programming.

Coverage probabilities are best calculated by computer. This necessitates the choice of a programming language and programming environment. Statisticians are generally familiar with one or more statistical programming languages such as SAS, R and so on. Such languages are necessary productivity tools due to their significant data handling capabilities as well as their statistical methods. They are indispensable to the statistician. However, they are not as useful as a language for describing algorithms. Small "bookkeeping" matters often obscure the algorithm or method to be calculated. This tutorial uses Mathematica as a language to describe the computation of coverage probabilities. With a little additional effort, one can produce graphs of coverage probabilities as well as dynamic demonstrations that use a slider to illustrate the effect of the sample size on the graph. The Wolfram Demonstrations Project website contains numerous Demonstrations involving a wide variety of topics. One such Demonstration provided by Heiner and Wagon [2] involves coverage probabilities for a population proportion using a Wald approach as well as a Bayesian approach. This article takes a different approach than Heiner and Wagon.

We illustrate the idea of coverage (and hence confidence) with several examples.

Section 2 describes two asymptotically justified confidence intervals for estimating a population proportion based on the binomial distribution. The first confidence interval is a simple hand calculation interval contained in many textbooks. We present a step-by-step algorithm for computing the coverage probability for one specific value of the population parameter. We stress clarity of computation rather than efficiency. The approach is adequate for a population described by a discrete distribution with a finite number of possible values. We then compute the coverage probability using a much higher-level function, `Probability`, to automatically compute the probability associated with an inequality. We also use `Probability` for subsequent calculations. We produce a typical graph of coverage probabilities found in some textbooks. The second confidence interval for a population proportion (again based on the binomial distribution) is more complicated but has gained popularity. Naturally, it will be seen that coverage probabilities are generally higher than the level of confidence when approximations are used to create a confidence interval. This is illustrated in the examples below.

Section 3 presents an asymptotically justified confidence interval for the mean of a population described by a Poisson distribution. The Poisson distribution has infinitely many possible observable values. The function `Probability` used to evaluate coverage probabilities automatically takes this into account.

Section 4 presents a graph of coverage probabilities based on an asymptotically justified confidence interval for estimating a population proportion based on the negative binomial distribution.

Section 5 presents a summary.

■ 2. A Population Proportion and the Binomial Distribution

□ The Simplest Confidence Interval

A population has a proportion p of members with a given characteristic. In order to estimate p, one randomly selects n members of the population with replacement, say $X_1, \ldots, X_n$, where the X_i are independent and identically distributed random variables, each with a Bernoulli distribution with parameter p. If X is the number of members in the random sample possessing the target characteristic, that is, $X = X_1 + X_2 + \ldots + X_n$, then X has a binomial distribution with parameters n and p. The sample proportion of members with the characteristic is $\hat{p} = X / n$. Two large sample confidence intervals for p are typically given. We start with the simplest. A large sample confidence interval of size $1 - \alpha$ for p is given by

$$\hat{p} \pm z_{\alpha/2} \sqrt{\hat{p}(1 - \hat{p}) / n} , \tag{1}$$

where $z_{\alpha/2}$ is the upper $\alpha / 2$ part of the standard normal distribution.

Let $\mathrm{se}(\hat{p}) = \sqrt{\hat{p}(1 - \hat{p}) / n}$, the standard error of $\hat{p}$. So, we may shorten (1) by writing it as

$$\hat{p} \pm z_{\alpha/2} \, \mathrm{se}\,(\hat{p}). \tag{2}$$

One can find the confidence interval in expression (2) in virtually any statistics book; in particular, see Devore and Berk [3, p. 396]. Also, coverage probabilities for this confidence interval are described in Brown, Cai and DasGupta [4]. The derivation of the interval leads one to believe that the level of confidence is $1 - \alpha$. However, two approximations are used to derive the interval in expression (2). One approximation uses the central limit theorem. A second approximation uses an estimated variance for the sampling distribution of the sample proportion p. We want to compute the actual coverage probability for any possible value of the true population proportion p. The coverage probability is

$$\mathrm{coverage}(p) = P\big(\hat{p} - z_{\alpha/2} \, \mathrm{se}\,(\hat{p}) < p < \hat{p} + z_{\alpha/2} \, \mathrm{se}\,(\hat{p}); \, p\big), \tag{3}$$

where $\hat{p} = X / n$. Books are sometimes vague about whether or not to include the endpoints in the inequality. We exclude the endpoints in order to be consistent with typical hypothesis testing methods.

The definition of coverage confuses many students. For a given value of p with $0 < p < 1$, one must determine the set values of X satisfying the inequality in expression

(3) and compute the probability of observing such values of *X*. We will describe how to determine the set of values and then compute their probability. Once we know what is actually being computed, we will move on to higher-level functions that perform the computations automatically.

We use an example with *n* = 10. A plot will show how bad the approximation can be and also displays the output of each step of the algorithm. We will compute the coverage probability for *p* = 0.5. The input and output are presented in a conversational style with some editorial comments along the way.

We wish to determine the upper 2.5 percentage value from the standard normal distribution. The variable `zc` is often called a critical value for the standard normal distribution. The result will be a floating-point number, which restricts the accuracy and precision of all calculations that use it; the result of this calculation is a floating-point number.

```
In[1]:= zc = Quantile[NormalDistribution[0, 1], 0.975]
```

```
Out[1]= 1.95996
```

Define `se`.

```
In[2]:= se = Sqrt[1/n x / n (1 - x / n)];
```

Here is the confidence interval inequality for sample size 10 and general `x`.

```
In[3]:= inequality[x_] = (x / n - zc se < p < x / n + zc se /. n → 10)
```

$$Out[3]= \frac{x}{10} - 0.195996 \sqrt{\left(1 - \frac{x}{10}\right) x} < p < \frac{x}{10} + 0.195996 \sqrt{\left(1 - \frac{x}{10}\right) x}$$

The support of the random variable *X* is the set of values *x* for which the probability mass function is positive. They also represent the observable values of *X* for a discrete random variable. We represent the support of *X* with the programming variable `xs`.

```
In[4]:= xs = Range[0, 10]
```

```
Out[4]= {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
```

This tests whether the inequality is true for each value of *X* and probability 0.5.

```
In[5]:= TF = inequality /@ xs /. p → 0.5
```

```
Out[5]= {False, False, False, True, True,
         True, True, True, False, False, False}
```

These are the positions that yield `True`; `Flatten` eliminates one level of parentheses. We wish to compute the probabilities of X at those positions and sum them.

```
In[6]:= positionsTrue = Flatten[ Position[TF, True] ]
```

```
Out[6]= {4, 5, 6, 7, 8}
```

These are the appropriate values of the variable X.

```
In[7]:= XTrue = xs[[ positionsTrue ]]
```

```
Out[7]= {3, 4, 5, 6, 7}
```

Now one computes the probabilities for the individual values of X satisfying the inequality.

```
In[8]:= PDF[BinomialDistribution[n, p], XTrue] /. {n → 10, p → 0.5}
```

```
Out[8]= {0.117188, 0.205078, 0.246094, 0.205078, 0.117188}
```

Finally, the values of the individual probabilities are summed to create the actual coverage probability for $p = 0.5$.

```
In[9]:= Total@%
```

```
Out[9]= 0.890625
```

The steps have been broken down so that students can easily understand what is needed. A large sample justification leads us to believe that this number should be about 0.95. The coverage probability is about 0.89 rather than 0.95.

Here is a much more transparent manner in which to compute the coverage probability. We may use a system function for evaluating the probability of expressions of a random variable. Apparently, the system function automatically tests each possible value of the random variable to determine the ones that satisfy the inequality. (This works quite well for a discrete random variable with a finite number of observable values.) The relevant probabilities are then summed. This approach is not efficient in cases with infinitely many observable values of a random variable. However, it is straightforward and easy for a student to understand. We evaluate the probability of an expression involving the binomial random variable. The expression of the binomial random variable is the confidence interval inequality.

Let us define a function that constructs the inequality more explicitly.

```
In[10]:= binomialInequality1[x_, n_, zc_, p_] :=
         With[
            {se = √(1/n x / n (1 - x / n))},
            x / n - zc se < p < x / n + zc se
         ]
```

Define the function that computes the coverage.

```
In[11]:= coverage1[n_, zc_, p_] := Module[
            {X},
            Probability[
              binomialInequality1[X, n, zc, p],
              X ≈ BinomialDistribution[n, p]
            ]
         ]
```

We now plot the coverage probabilities for a range of values of *p* in Figure 1 below. We also create a horizontal line at a level of 0.95 for comparison purposes. The graph is symmetric due to properties of the binomial distribution and the large sample approximation involved in the confidence interval justification.

```
In[12]:= ListPlot[
            Table[{p, coverage1[10, zc, p]}, {p, 0, 1, 0.002}],
            AxesLabel → {p, "coverage"},
            Prolog → {Red, Line[{{0, 0.95}, {1, 0.95}}]}
         ]
```

Out[12]=
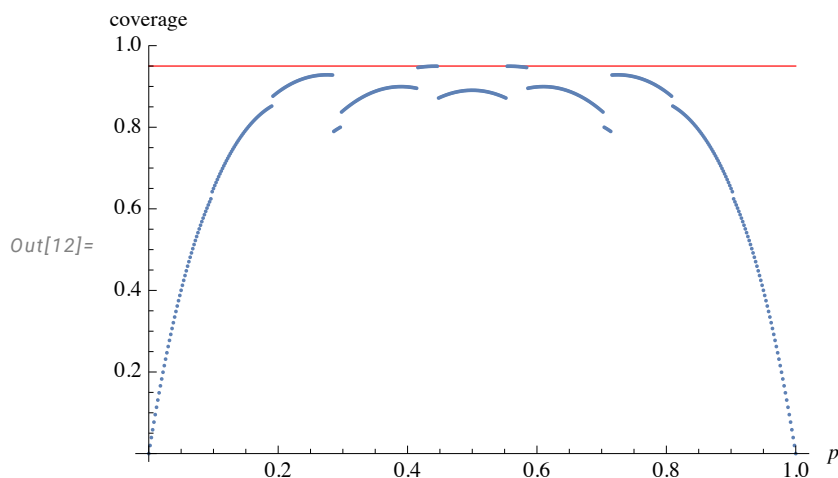
▲ **Figure 1.** Coverage plot for first binomial confidence interval, *n* = 10.

Examining Figure 1 indicates several points. First, the coverage probabilities are in general not equal to the nominal level of confidence—namely .95. Moreover, coverage probabilities near $p = 0$ and $p = 1$ are effectively zero. Finally, the coverage probability function is discontinuous. All this with a minimum level of programming. In fact, the programming statements presented are simply a good description of the algorithm.

More is available. We wish to be able to change the plot by varying the sample size with a slider. A dynamic demonstration can easily be created with the `Manipulate` function. The manipulate variable is the sample size n, which you can vary with a slider from 5 to 100.

The graph is in Figure 2. The computer processing time increases with the value of the sample size n because the inequality must be tested for each possible value of X. The initial sample size is $n = 50$.

```
In[13]:= Manipulate[
         ListPlot[
          Table[{p, coverage1[n, zc, p]}, {p, 0, 1, 0.005}],
          PlotRange → {0, 1},
          AxesLabel → {p, "coverage"},
          Prolog → {Red, Line[{{0, 0.95}, {200, 0.95}}]}]
         ],
         {{n, 50}, 5, 100, 1, Appearance → "Labeled"},
         SaveDefinitions → True,
         TrackedSymbols :> {n}
        ]
```
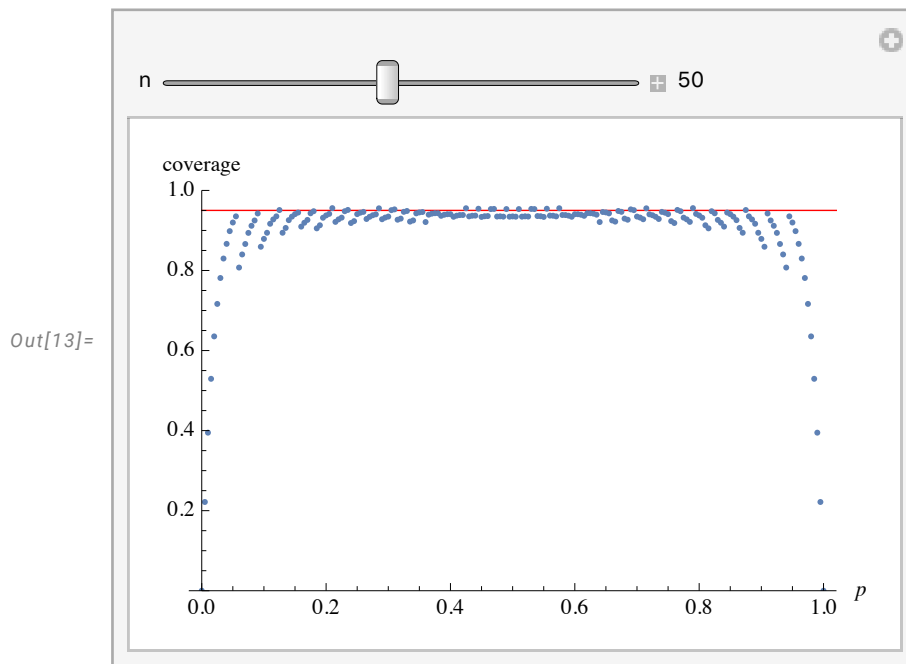

▲ **Figure 2.** Coverage plot as a function of sample size.

A larger sample size improves the coverage probabilities as one expects. After all, the confidence interval formula is justified by a large sample argument. However, it is very clear that the coverage probability is small when p is close to either 0 or 1 even with $n = 50$. For some sample sizes it is even more obvious that this function contains discontinuities.

▫ A Better Confidence Interval for a Population Proportion

This subsection presents coverage probabilities for an improved confidence interval for a population proportion. The improvement makes coverage probabilities generally larger.

Devore and Berk [3, p. 395] give a better large sample confidence interval for a population proportion. This interval was previously presented in Agresti and Coull [5]. Based on the same assumptions as expression (1), a sample confidence interval of size $1 - \alpha$ for a population proportion p is given by

$$\frac{\hat{p} + \left(z_{\alpha/2}^2 / 2n\right)}{1 + z_{\alpha/2}^2 / n} \pm z_{\alpha/2} \times \sqrt{\hat{p}(1 - \hat{p})/n + z_{\alpha/2}^2 / (4n^2)} \bigg/ \left(1 + z_{\alpha/2}^2 / n\right). \tag{4}$$

This confidence interval is based on solving the following inequality for p:

$$-z_{\alpha/2} < \left(\hat{p} - p\right) \bigg/ \sqrt{p(1 - p)/n} < z_{\alpha/2}. \tag{5}$$

This defines the new inequality accordingly.

```
In[14]:= binomialInequality2[x_, n_, zc_, p_] :=
    With[
        {ptilde = (x / n + zc^2 / (2 n)) / (1 + (zc^2 / n)),
        s = Sqrt[x / n (1 - x / n) / n + zc^2 / (4 n^2)] / (1 + zc^2 / n)},
        ptilde - zc s < p < ptilde + zc s
    ]
```
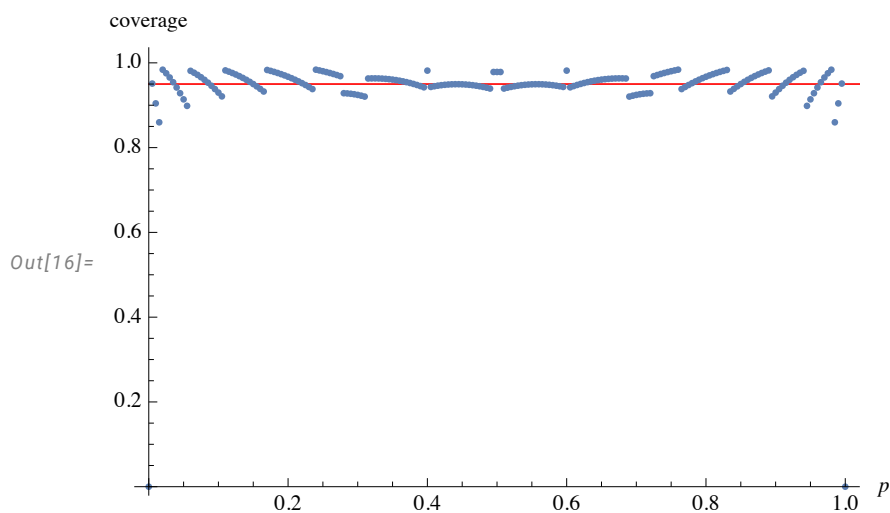
Just as with the previous kind of inequality, define `coverage2`.

```
In[15]:= coverage2[n_, zc_, p_] := Module[
    {X},
    Probability[
        binomialInequality2[X, n, zc, p],
        X ≈ BinomialDistribution[n, p]
    ]
]
```

Figure 3 is the corresponding plot, again with $n = 10$.

```
In[16]:= ListPlot[
         Table[{p, coverage2[10, zc, p]}, {p, 0, 1, .005}],
         AxesLabel → {p, "coverage"},
         Prolog → {Red, Line[{{0, 0.95}, {200, 0.95}}]}
        ]
```

Out[16]=

▲ **Figure 3.** Coverage plot for the better confidence interval, $n = 10$.

The inequality in (5) is supposed to have a probability of approximately $1 - \alpha$ before sampling the population. We can of course compute the true probability with respect to the correct binomial distribution. The Mathematica code follows along with a dynamic graph in Figure 4.

```
In[17]:= Manipulate[
        ListPlot[
         Table[{p, coverage2[n, zc, p]}, {p, 0, 1, .005}],
         PlotRange → {0, 1},
         AxesLabel → {p, "coverage"},
         Prolog → {Red, Line[{{0, 0.95}, {1, 0.95}}]}]
        ],
        {{n, 50}, 5, 100, 1, Appearance → "Labeled"},
        SaveDefinitions → True,
        TrackedSymbols :> {n}
       ]
```

Out[17]=
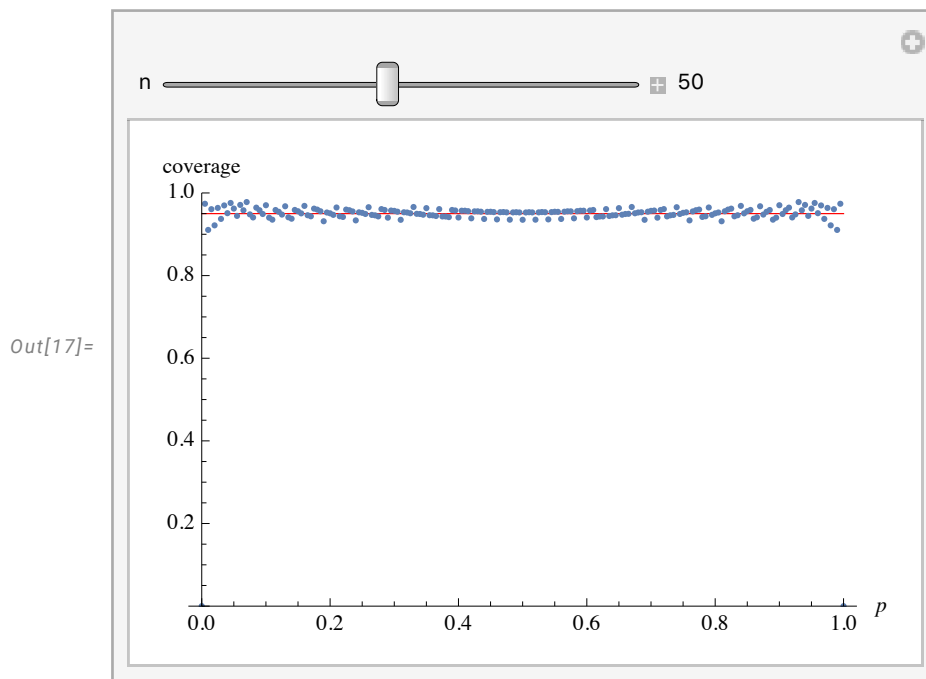
▲ **Figure 4.** Coverage probabilities for the superior asymptotic confidence interval for a population proportion.

Figure 5 contains the code and plot for the dynamic version of the plot. This plot allows for an easy comparison of the coverage probabilities for the two types of confidence intervals.

```
In[18]:= Manipulate[
  ListPlot[
    Transpose@
     Table[{{p, coverage1[n, zc, p]}, {p, coverage2[n, zc, p]}},
       {p, 0, 1, 0.005}],
     PlotRange → {0, 1},
     AxesLabel → {p, "coverage"},
     Prolog → {Red, Line[{{0, 0.95}, {1, 0.95}}]}
   ],
   {{n, 50}, 5, 100, 1, Appearance → "Labeled"},
   SaveDefinitions → True,
   TrackedSymbols :> {n}
 ]
```
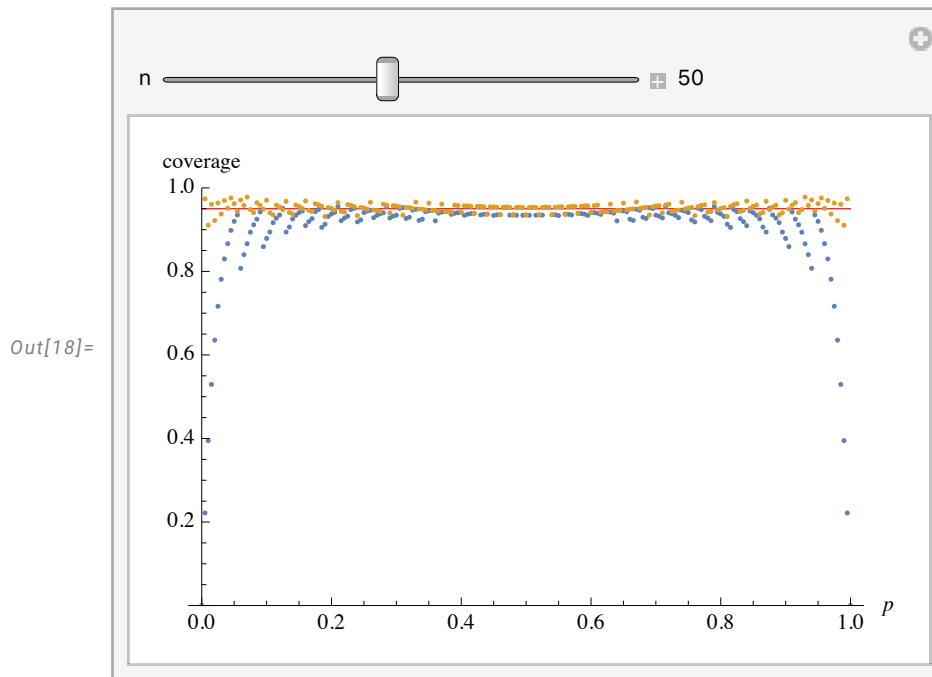

▲ **Figure 5.** A comparison of coverage probabilities for the two binomial intervals.

The coverage probabilities for this improved confidence interval for a population proportion are indeed superior to the simpler interval. In particular, the coverage probabilities are quite large when p is close to 0 or 1. One can see this even with a sample size of $n = 10$, for which the large sample approximation is not appropriate. The difference in coverage probabilities with the simple interval (displayed in Figure 2) and this improved interval is striking.

◼ 3. The Mean of the Poisson Distribution

We now turn our attention to the Poisson distribution.

The book by Devore and Berk [3, p. 400] presents a homework exercise for determining a confidence interval of size $1 - \alpha$ for the mean θ of a population described by a Poisson distribution. Let $X_1, X_2, \ldots, X_n$, where the X_i are independent and identically distributed with a Poisson distribution with parameter (mean) of θ. Ideally, we must solve the inequality

$$-z_{\alpha/2} < \left(\overline{X} - \theta\right) / \sqrt{\theta/n} < z_{\alpha/2} \tag{6}$$

to obtain the desired confidence interval. However, if we have a large enough sample, we may replace the true standard error in the denominator with its estimate. Again, this produces a less than ideal result.

The resulting simple confidence interval of approximate size $1 - \alpha$ for the mean θ is given by

$$\overline{X} - z_{\alpha/2} \sqrt{\overline{X}/n} < \theta < \overline{X} + z_{\alpha/2} \sqrt{\overline{X}/n}, \tag{7}$$

which has an approximate level of confidence of $1 - \alpha$. The parameter θ in the denominator was replaced by the sample mean. Figure 6 contains the code and graph for the coverage probabilities. We use $n = 50$. We let $\overline{X} = Y/50$ where $Y = \sum_{i=1}^{50} X_i$ has a Poisson distribution with a mean of $50\,\theta$. In principle, the inequality must be tested for each of the infinitely many possible values of Y. Coverage probabilities are evaluated at a discrete set of points in order to save computational time.

```mathematica
In[19]:= Module[
        {n = 50, X, θ, zc},
        zc = Quantile[NormalDistribution[0, 1], .975];
        ListPlot[
         θ = Range[0, 3, .01];
         {#, Probability[
             Evaluate[
               X / n - zc / n √X < # < X / n + zc / n √X],
               X ≈ PoissonDistribution[n #] ]} & /@ θ,
         AxesLabel → {"θ", "coverage"},
         PlotRange → {{0, 3}, {0, 1}},
         Prolog → {Red, Line[{{0, 0.95}, {3, 0.95}}]}]
        ]
       ]
```

Out[19]=
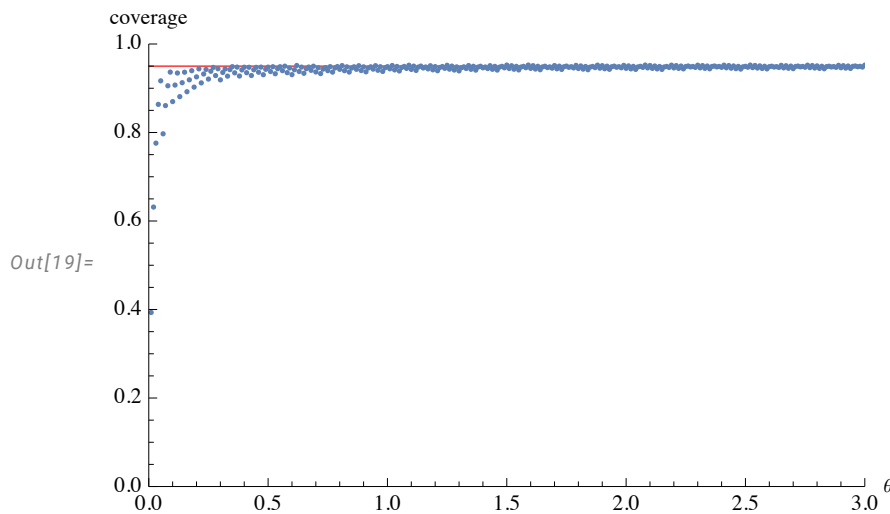
▲ **Figure 6.** Coverage probabilities for the confidence interval for the Poisson mean $n = 50$.

Unless θ is close to zero, this large sample approximation is quite good for $n = 50$, which is easily seen in Figure 6. Given the two approximations used, it is not surprising that the coverage probability is small when θ is close to zero.

■ 4. The Population Proportion and the Negative Binomial Distribution

This section addresses the situation of estimating a population proportion when the negative binomial distribution is appropriate.

Let $X_1, X_2, \ldots, X_n$, where the X_i are independent and identically distributed with a geometric distribution with parameter p. It is well known that $X = \sum_{i=1}^{n} X_i$ has a negative binomial distribution with parameters n and p, (see Kinney [6], p. 127). Consequently, we use the negative binomial distribution for estimating a population proportion. There are many ways to define the negative binomial distribution. We use the version described in Kinney [6, p. 125]. Conduct independent success/failure trials, each with a probability of success p. Let X be the total number of trials needed to obtain n successes. The probability mass function for X is given by

$$f(x) = \binom{x - 1}{n - 1} p^n (1 - p)^{x-n}, \, x = n, n + 1, \ldots, \tag{8}$$

where $0 < p < 1$.

Some authors count the number of trials before the nth success. Other authors count the number of failures before the nth success. There are other possibilities still. Mathematica uses Y, the number of failures before the nth success. Consequently, $Y + n = X$.

Casella and Berger [1, p. 496] describe large sample confidence intervals based on maximum likelihood. It is easily shown that the maximum likelihood estimator for p is $\hat{p} = n/X$. Moreover, the asymptotic variance of this estimator is the reciprocal of the Fisher information, $\mathrm{Var}(\hat{p}) = p^2(1-p)/n$. Fisher information is described in Casella and Berger [1, p. 388]. This variance expression is not useful for creating a confidence interval for p since it depends on p. So, we estimate the large sample variance by replacing p with $\hat{p}$. This leads to the large sample confidence interval:

$$1 - \alpha \approx P\left(\hat{p} - z_{\frac{\alpha}{2}}\ \sqrt{\hat{p}^2(1-\hat{p})\big/ n}\ < p < \hat{p} + z_{\frac{\alpha}{2}}\ \sqrt{\hat{p}^2(1-\hat{p})\big/ n}\right). \tag{9}$$

In order to conveniently perform the calculations, we note that $\hat{p} = n/X = n/(Y+n)$. We evaluate the coverage probability for $0.2 < p < 0.99$ in steps of 0.01. We based the calculations on $n = 20$. The calculation can take some time depending on the computer. When p is small, values of X or Y are extremely unlikely. This makes the internal algorithm take quite a while. We can help speed up the calculations by using `NProbability` rather than the symbolic `Probability`. The speedup occurs by reducing the required number of digits in calculations. Even so, this calculation takes some time (about four minutes on the author's computer). A graph of the coverage probabilities is contained in Figure 7.

```
In[20]:= Module[

        {zc, n, se, p, X},
        zc = Quantile[NormalDistribution[0, 1], .975];
        n = 20;

        se = n/(X + n) Sqrt[(1 - n/(X+n))/n] ;

        p = Range[.2, 1, 0.01];
        ListPlot[

          {#, NProbability[

              n/(X + n) - zc se < # < n/(X + n) + zc se,

              X ≈ NegativeBinomialDistribution[n, #]]} & /@ p,

          AxesLabel → {Style["p", Italic], "coverage"},
          PlotRange → {{0, 1}, {0, 1}},
          Prolog → {Red, Line[{{0, 0.95}, {1, 0.95}}]}]

        ]

      ]
```

Out[20]=
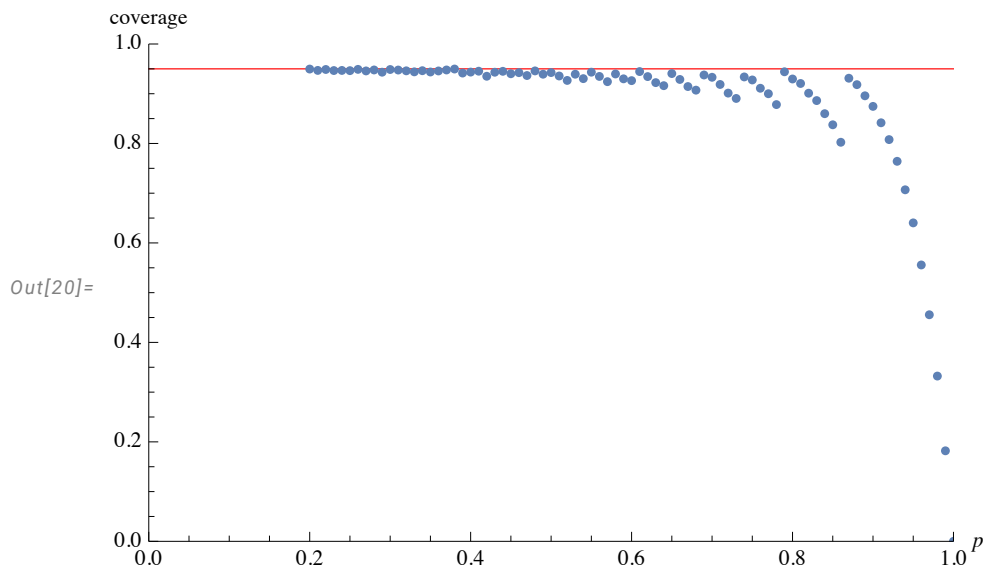
▲ **Figure 7.** Coverage probabilities for the confidence interval for a population proportion on the negative binomial distribution, $n = 20$.

We see from Figure 7 that the approximation is quite good for values of p close to 0.2. We infer that the approximation is also quite good if p is close to 0. The approximation generally gets worse as p increases (though not monotonically). A large sample approximation was used. Also, an approximate standard error was used. One sees that the coverage probability is essentially zero when p is close to 1.

■ 5. Summary

Large sample confidence intervals are often quite easy to derive. This is particularly true when using an estimate for the standard error of an estimator. However, the actual probability of surrounding the parameter value (coverage) can be quite different from the nominal value. It is helpful to graph the coverage probabilities to see this. Mathematica is particularly useful in performing these calculations and providing a language for describing the algorithms.

■ Acknowledgments

■ References

[1] G. Casella and R. Berger, *Statistical Inference,* 2nd ed., United States: Brooks/Cole Cengage Learning, 2002.

[2] K. Heiner and S. Wagon. "Wald and Bayesian Confidence Intervals" from the Wolfram Demonstrations Project—A Wolfram Web Resource. www.demonstrations.wolfram.com/WaldAndBayesianConfidenceIntervals.

[3] J. Devore and K. Berk, *Modern Mathematical Statistics with Applications*, 2nd ed., New York: Springer, 2012.

[4] L. D. Brown, T. T. Cai and A. DasGupta, "Confidence Intervals for a Binomial Proportion and Asymptotic Expansions," The *Annals of Statistics*, **30**(1), 2002 pp. 160–201. www.jstor.org/stable/2700007.

[5] A. Agresti and B. A. Coull, "Approximate Is Better than 'Exact' for Interval Estimation of Binomial Proportions,", *The American Statistician*, **52**(2), 110-126. http://doi.org/10.1080/00031305.1998.10480550.

[6] J. Kinney, *Probability: An Introduction with Statistical Applications*, New York: John Wiley and Sons, 1997.

About the Author

Peyton Cook earned a B.A. in Psychology, B.S. in Mathematics, and an M.S. and Ph.D. in Statistics. He is an Associate Professor at The University of Tulsa.

Peyton Cook
Department of Mathematics
The University of Tulsa
800 Tucker Drive
Tulsa, Oklahoma 74104
pcook@utulsa.edu