

Lower System Reliability Bounds from Binary Failure Data Using Bootstrapping

LAWRENCE M. LEEMIS

The College of William & Mary, Williamsburg, Virginia 23187

Binary failure data are collected for each of the independent components in a coherent system. Bootstrapping is used to determine a $(1 - \alpha)100\%$ lower confidence bound on the system reliability. When a component with perfect test results is encountered, a beta prior distribution is used to avoid an overly optimistic lower bound.

Key Words: Beta distribution, Binomial confidence interval, Coherent system, Computer algebra system.

WE CONSIDER the problem of determining a $(1 - \alpha)100\%$ lower confidence bound on the system reliability for a coherent system of k components using the failure data (y_i, n_i) , where y_i is the number of components of type i that pass the test and n_i is the number of components of type i on test, $i = 1, 2, \dots, k$. We assume throughout that the components fail independently, e.g., no common-cause failures. The outline of the article is as follows. We begin with the case of a single ($k = 1$) component system where n components are placed on a test and y components pass the test. The Clopper-Pearson lower bound is used to provide a lower bound on the reliability. This model is then generalized to the case of multiple ($k > 1$) components. Bootstrapping is used to estimate the lower confidence bound on system reliability. We then address a weakness in the bootstrapping approach—the fact that the sample size is moot in the case of perfect test results, e.g., when $y_i = n_i$ for some i . This weakness is overcome by using a beta prior distribution to model the component reliability before performing the bootstrapping. Two subsections consider methods for estimating the parameters in the beta prior distribution for components with perfect test results. The first subsection considers the case when previous test results are available, and the second subsection considers the case when no previous test results are available. A simulation study compares various algorithms for

calculating a lower confidence bound on the system reliability. The last section contains conclusions.

Single-Component Systems

Single-component systems are considered first because (1) there are known approximate confidence intervals for the lower reliability confidence bound and (2) these intervals will be used later in the article to help determine the appropriate parameters for the beta distribution in the case where no prior test results exist on the component of interest.

Let n components be placed on test and let y of these components pass the test. Under the assumption that the test values (1 for pass, 0 for failure) X_1, X_2, \dots, X_n are independent and identically distributed Bernoulli random variables with unknown parameter p , $Y = \sum_{i=1}^n X_i$ is a binomial random variable with parameters n and p . The maximum likelihood estimator for p is $\hat{p} = Y/n$, which is unbiased and consistent. The interest here is in a lower confidence bound for the reliability p .

There is a wide literature on confidence intervals of this type because a confidence interval on a proportion is of interest on anything from a political poll to consumer preference. Vollset (1993) compares 13 confidence intervals and Newcombe (2001) compares 7 confidence intervals. Rather than fine tuning these intervals, as has been suggested by many authors, we have settled on using the Clopper-Pearson (CP) exact interval even though Newcombe (2001, page 201) points out that its status as a gold standard has been disputed recently because the method is conserva-

Dr. Leemis is a Professor in the Department of Mathematics. His email address is leemis@math.wm.edu.

tive, i.e., the actual coverage is greater than or equal to the stated coverage (see Agresti and Coull, 1998, for details).

Let $p_L < p < p_U$ be an exact (Blyth 1986) CP two-sided confidence interval for p , where p_L and p_U are functions of the sample size n , the number of successes y , and the stated coverage of the interval, $1 - \alpha$. This is an approximate confidence interval due to the discrete nature of the binomial distribution. For $y = 1, 2, \dots, n - 1$, the lower limit p_L satisfies (see, for example, Agresti and Coull, 1998)

$$\sum_{k=y}^n \binom{n}{k} p_L^k (1 - p_L)^{n-k} = \alpha/2.$$

For $y = 1, 2, \dots, n - 1$, the upper limit p_U satisfies

$$\sum_{k=0}^y \binom{n}{k} p_U^k (1 - p_U)^{n-k} = \alpha/2.$$

As shown in Leemis and Trivedi (1996), these confidence-interval limits can be expressed in terms of quantiles of the F distribution:

$$1 + \frac{1}{yF_{2y, 2(n-y+1), 1-\alpha/2}} < p < \frac{1}{1 + \frac{1}{(y+1)F_{2(y+1), 2(n-y), \alpha/2}}},$$

where the third subscript on F refers to the right-hand tail probability.

Simply reallocating the probability α to the lower limit gives the following lower confidence bound for the reliability:

$$p_L = \frac{1}{1 + \frac{1}{yF_{2y, 2(n-y+1), 1-\alpha}}},$$

for $y = 1, 2, \dots, n - 1$. For the case of all failures ($y = 0$), the lower bound is, of course, $p_L = 0$. For the case of all passes ($y = n$), the lower bound is $p_L = \alpha^{1/n}$.

Example: CP Lower Confidence Interval Bounds

The following four sets of values for n and y give point estimates and 95% CP lower confidence interval bounds for the reliability:

- $n = 10, y = 7 \Rightarrow \hat{p} = 0.7, p_L = 0.393.$
- $n = 100, y = 97 \Rightarrow \hat{p} = 0.97, p_L = 0.924.$
- $n = 10, y = 10 \Rightarrow \hat{p} = 1.0, p_L = 0.741.$
- $n = 100, y = 100 \Rightarrow \hat{p} = 1.0, p_L = 0.970.$

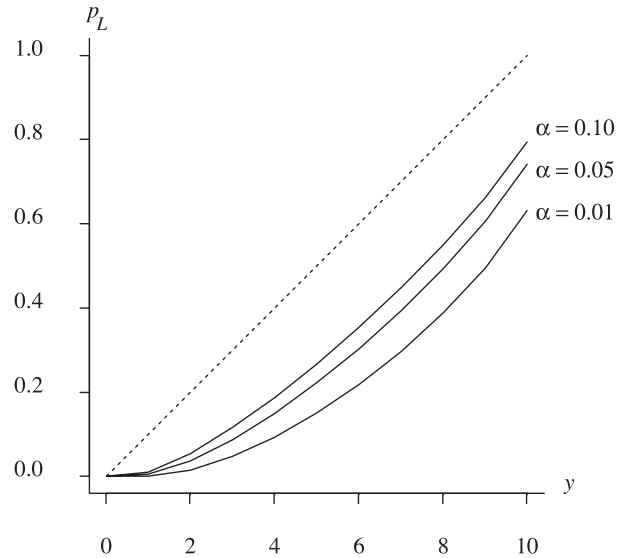


FIGURE 1. Point Estimate (dashed) and CP Lower Confidence Bounds (solid) When $n = 10$.

An S-Plus function named `confintlower` is given in Appendix 1, which can be used to calculate these lower confidence-interval bounds. Figure 1 is a plot of y vs. p_L when $n = 10$ for $\alpha = 0.10, 0.05, 0.01$, with the points connected with line segments. Figure 2 contains a similar plot for $n = 100$. The lower bounds are monotonic in y, n , and α .

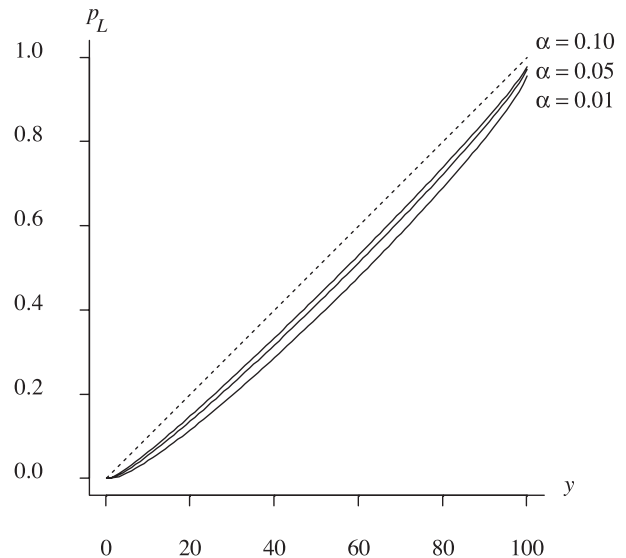


FIGURE 2. Point Estimate (dashed) and CP Lower Confidence Bounds (solid) When $n = 100$.

Multiple-Component Systems

A three-component ($k = 3$) series system is used as an example throughout this section, although the techniques described here apply to any coherent system of k independent components. The number of components tested and the number of passes for each type of component for the example are given in Table 1. The point estimate for the system reliability is

$$\frac{21}{23} \cdot \frac{27}{28} \cdot \frac{82}{84} = \frac{1107}{1288} \cong 0.8595.$$

The remainder of this section involves the use of bootstrapping (Efron and Tibshirani, 1993) to calculate a lower 95% confidence-interval bound. Other authors (e.g., Martin (1990) and Padgett and Tomlinson (2003)) have used bootstrapping for determining confidence limits. The approach used here differs conceptually from the standard bootstrap problem, where the standard error of a *single* unknown distribution is estimated by resampling iid data. In our setting, there are k *different* distributions (one for each component) and we resample component reliabilities and combine using the reliability function to yield the system reliability.

Bootstrapping resamples B of the systems, calculates the system reliability, then outputs the αB^{th} ordered system reliability. More specifically, for the three-component system of interest, the bootstrapping algorithm follows these steps.

- For the first component, the data set (21 ones and 2 zeros) is sampled with replacement 23 times.
- These values are summed and divided by 23, yielding a reliability estimate for the first component.
- The previous two steps are repeated for components 2 and 3.
- The product of the reliability estimates for the three components are multiplied (because the components are arranged in series and their failures are independent) to give a system-reliability estimate.

The above procedure is repeated B times. The B system reliability estimates are then sorted. Finally, the αB^{th} ordered system reliability is output, which is used as a lower bound on the system reliability.

The algorithm for estimating the $(1 - \alpha)100\%$ lower confidence-interval bound is given in Table 2, where \tilde{p}_i is a bootstrap estimate for the reliability of component i and z_j is a bootstrap estimate of the system reliability. The binomial distribution is appropriate because the resampling from the data set

TABLE 1. Failure Data for
a Three-Component Series System

Component number	$i = 1$	$i = 2$	$i = 3$
Number passing (y_i)	21	27	82
Number on test (n_i)	23	28	84

is performed with replacement. In the pseudocode in Table 2, indentation is used to indicate begin-end blocks. The returned value $z_{\alpha B}$ is the order statistic associated with the z_j 's generated in the outside loop. See Law and Kelton (2000) for handling the case when αB is not an integer.

This algorithm has been implemented in S-Plus as a function named `seriessystemboot`, which is given in Appendix 2. The first two arguments, `n` and `y`, are vectors of length k , and the third argument, `alpha`, is a real number between 0 and 1, e.g.,

```
seriessystemboot(c(23, 28, 84),
                 c(21, 27, 82), 0.05)
```

prints a point estimate and a 95% lower confidence-interval bound on the system reliability for the three-component series system considered in this section. After a call to `set.seed(3)` to set the random number seed, five calls to `seriessystemboot` yield the following estimates for p_L :

```
0.7426057 0.7482993 0.7456744
0.7486690 0.7453416.
```

The dispersion associated with these five estimates is due to the finite choice of B , i.e., $B = 10,000$.

Resampling error can be eliminated using the symbolic Maple-based APPL (Glen, Evans and Leemis, 2001). The APPL statements given in Appendix 3 utilize the `Product` and `Transform` procedures. This alternative approach to determining a lower 95% bootstrap confidence interval bound for the system reliability is equivalent to using an infinite value for B . Because \tilde{p}_i can assume any one of $n_i + 1$ values, there are a possible $24 \cdot 29 \cdot 85 = 59,160$ potential mass values for the random variable T determined by the `Product` procedure. Of these, only 6633 are distinct because the `Product` procedure combines repeated values. Because the random variable T from the APPL code plays an analogous role to the vector `z` from the bootstrap algorithm in Table 2, the lower 95% bootstrap confidence-interval

TABLE 2. Bootstrap Algorithm for Calculating a $(1 - \alpha)100\%$ Lower Confidence Bound for the Reliability of a k -Component Series System

for j from 1 to B	[resampling loop]
for i from 1 to k	[loop through components]
$\tilde{p}_i \leftarrow \text{Binomial}(n_i, y_i/n_i)/n_i$	[component i reliability]
$z_j \leftarrow \prod_{i=1}^k \tilde{p}_i$	[calculate system reliability]
sort \mathbf{z}	[sort the system reliability values]
return $z_{\alpha B}$	[return the estimate for the lower bound]

bound is the 0.05 fractile of the distribution of T , which is $p_L = 6723/9016 \cong 0.746$. This result using APPL is consistent with the standard resampling approach for finding the lower confidence-interval limit based on the five results presented earlier (one equals 6723/9016 exactly, two fall above 6723/9016, and two fall below 6723/9016).

How Well Does the Bootstrap Procedure Perform?

This question is difficult to address because there is no exact interval to compare with, even in the case of a single component. It is instructive, however, to isolate one component and compare the CP approach described in the previous section with bootstrapping. Arbitrarily choosing the second component with $n_2 = 28$ items on test, Figure 3 shows the CP lower confidence bound and the bootstrap lower confidence bound for $\alpha = 0.05$ and for $y_2 = 10, 11, \dots, 28$. The bootstrap lower confidence interval limit does not require any iteration because the value plotted is $q/28$, where q is the smallest integer that satisfies

$$\sum_{k=0}^q \binom{n_2}{k} \left(\frac{y_2}{28}\right)^k \left(1 - \left(\frac{y_2}{28}\right)\right)^{n_2-k} \geq \alpha,$$

i.e., q is the α -quantile of a binomial distribution with $n_2 = 28$ trials and probability of success $y_2/28$. Figure 3 shows that

- the bootstrap interval is more susceptible to the discrete nature of the binomial sampling scheme
- the CP interval is wider than the bootstrap interval.

Figure 3 also points out a glaring deficiency in the bootstrapping approach that was not revealed in the example in this section because all three of the system components had one or more failures during their life test. When component i has perfect test results (e.g., $y_i = n_i$), the sample size becomes irrelevant. Thus, a test where two components out of two

pass the test is equivalent to one where 100 components out of 100 pass the test from the perspective of the bootstrapping algorithm. This is clearly unacceptable. The next section gives a modification to the bootstrapping approach that adjusts for these perfect tests.

Perfect Component Test Results

The problem created by perfect component test results is likely to occur for components and systems with moderate to high reliability. As suggested by Chick (2001) and Martz and Waller (1982, pp. 265–266), a $\text{beta}(\alpha_1, \alpha_2)$ prior distribution can be placed on the component reliability. The beta distribution is a logical choice for a prior distribution of the component reliability due to (1) the flexibility in

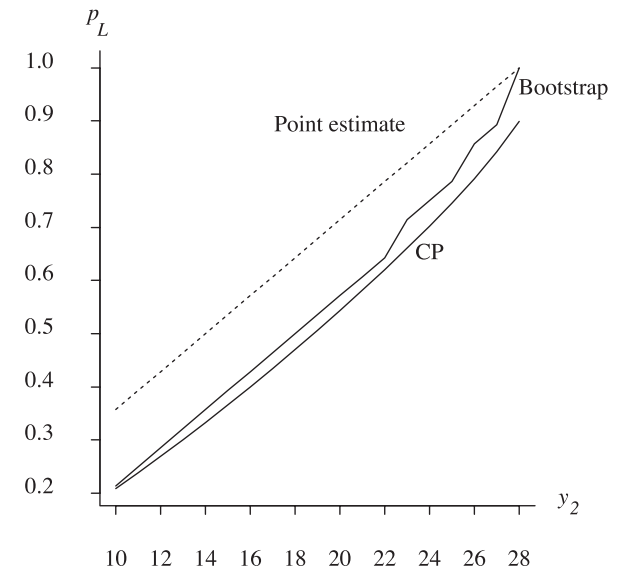


FIGURE 3. Point Estimate (dashed), CP Lower 95% Confidence Bound, and Bootstrap Lower 95% Confidence Bound for the Reliability of Component 2 Based on a Sample of Size $n_2 = 28$.

TABLE 3. Bootstrap Algorithm for Calculating a $(1 - \alpha)100\%$ Lower Confidence Bound for the Reliability of a k -Component Series System When Some Components Have Perfect Test Results

for j from 1 to B	[resampling loop]
for i from 1 to k	[loop through components]
if $(y_i = n_i) \tilde{p}_i \leftarrow \text{Beta}(\alpha_{1i} + y_i, \alpha_{2i})$	[component i reliability: perfect test]
else $\tilde{p}_i \leftarrow \text{Binomial}(n_i, y_i/n_i)/n_i$	[component i reliability: failure(s) occur]
$z_j \leftarrow \prod_{i=1}^k \tilde{p}_i$	[calculate system reliability]
sort \mathbf{z}	[sort the system reliability values]
return $z_{\alpha B}$	[return the estimate for the lower bound]

the shape of its probability density function, (2) its $(0, 1)$ support, and (3) its analytically tractable conjugate posterior distribution. Determining the values of the parameters α_1 and α_2 is a problem that will be addressed in the following two subsections.

The beta distribution has probability density function

$$f(x) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x^{\alpha_1-1} (1-x)^{\alpha_2-1}, \quad 0 < x < 1,$$

where α_1 and α_2 are positive shape parameters. This is the standard parameterization, although Martz and Waller (1982) use a slightly different form. The mean of a $\text{beta}(\alpha_1, \alpha_2)$ random variable is

$$\mu = E[X] = \frac{\alpha_1}{\alpha_1 + \alpha_2},$$

and the variance is

$$\sigma^2 = V[X] = \frac{\alpha_1 \alpha_2}{(\alpha_1 + \alpha_2)^2 (\alpha_1 + \alpha_2 + 1)}.$$

If the prior distribution of a reliability $P \sim \text{beta}(\alpha_1, \alpha_2)$ and the sampling is binomial (as it is in our case), then the posterior distribution of P is $\text{beta}(\alpha_1 + y, \alpha_2 + n - y)$, where n is the number of components on test and y is the number of passes. The difficulty in our case is in determining the appropriate values for α_1 and α_2 . For the time being, we will proceed as if we know the values of α_1 and α_2 and give an algorithm for finding the lower reliability confidence bound, p_L . Estimating α_1 and α_2 will be addressed subsequently.

A lower reliability confidence bound p_L can be determined by generating a bootstrap beta random variate (rather than a binomial) when the component test results are perfect. An algorithm for determining p_L for a k -component series system using B resamplings with some or all components having perfect tests is given Table 3. If component i has perfect test

results (i.e., $y_i = n_i$) then the analyst must define the prior beta parameters α_{1i} and α_{2i} .

Example: Three-Component Series System

Table 4 is identical to Table 1 except that component 2 now has perfect (28/28) test results. The point estimate for the system reliability increases to

$$\frac{21}{23} \cdot \frac{28}{28} \cdot \frac{82}{84} = \frac{41}{46} \cong 0.8913.$$

Thus, the effect of the one additional component that passed the test increases the system reliability estimate from approximately 0.86 to approximately 0.89. This increase should be reflected in an appropriate increase in the lower confidence limit, p_L .

This algorithm has been implemented as the S-Plus function `seriesystembayesboot` given in Appendix 4 (the number of bootstrap replications $B = 10,000$ and values of the beta parameters $\alpha_{12} = 1$ and $\alpha_{22} = 1$ are arbitrary). As before, `n` and `y` are vectors of length k and `alpha` is a real number between 0 and 1, e.g.,

```
seriesystembayesboot(c(23, 28, 84),
                    c(21, 28, 82), 0.05)
```

prints a point estimate and a 95% lower confidence-interval bound on the system reliability for the three-component series system. After a call to `set.seed(3)` to set the random number seed, five

TABLE 4. Failure Data for a Three-Component Series System

Component number	$i = 1$	$i = 2$	$i = 3$
Number passing (y_i)	21	28	82
Number on test (n_i)	23	28	84

calls to `seriessystembayesboot` yield the following values for p_L :

0.7474437 0.7484738 0.7492014
0.7484301 0.7495972.

With the choice $\alpha_1 = \alpha_2 = 1$, the increase of approximately 0.03 in the point estimate of the system reliability from the previous example results in only a tiny increase in the lower confidence interval limits. This is clearly unacceptable. What happened? The arbitrary choice of $\alpha_1 = 1$ and $\alpha_2 = 1$ has resulted in a uniform prior distribution, which is an overly pessimistic assessment of the reliability of component 2, particularly in light of the perfect test results.

What choice would make more sense? It is important to skew the probability density function of the beta prior distribution so that its mean is greater than 1/2, or, equivalently, choose $\alpha_2 < \alpha_1$. There are four different shapes of the probability density function associated with the choice of parameters that satisfy the constraint $\alpha_2 < \alpha_1$. Most important is the value of the probability density function near $f(1)$ because these are the particular reliability values of interest. The following four cases demark various features of the probability density function.

- $f(0) = 0$ and $f(1) = 0$ when $1 < \alpha_2 < \alpha_1$ (Case I).
- $f(1)$ is finite when $1 = \alpha_2 < \alpha_1$ (Case II).
- a vertical asymptote at $x = 1$ and $f(0) > 0$ when $\alpha_2 < 1 = \alpha_1$ (Case III).
- a vertical asymptote at $x = 1$ and $f(0) = 0$ when $\alpha_2 < 1 < \alpha_1$ (Case IV).

We have disregarded the case $\alpha_2 < \alpha_1 < 1$ because this results in a vertical asymptote at both 0 and 1, which is inconsistent with the probability density function of a high-reliability component. The most intuitively appealing of the four cases listed above is the fourth case, $\alpha_2 < 1 < \alpha_1$, because this minimizes the probability of generating a small beta variate (since $f(0) = 0$) and pushes as much of the probability near 1 as possible due to the vertical asymptote near 1.

Table 5 gives means of the beta prior distribution and lower confidence interval bounds for several combinations of α_1 and α_2 satisfying the constraint $\alpha_2 < \alpha_1$. The lower bounds are determined by taking the sample median of five runs of `seriessystembayesboot` with $B = 10,000$ resampled series systems per run. The subscript on the lower bound indicates which of the shapes in the list given above

TABLE 5. Prior Beta Distribution Mean and Lower 95% Confidence Interval Limit Estimate for the System Reliability

α_1	α_2	0.1	1	10
1	0.909/0.779 _{III}	—	—	—
10	0.990/0.783 _{IV}	0.909/0.759 _{II}	—	—
100	0.999/0.783 _{IV}	0.990/0.779 _{II}	0.909/0.725 _I	—

is represented. The value of the lower bound is quite sensitive to the choices of α_1 and α_2 . There are many (α_1, α_2) pairs that yield a reasonable lower bound.

The following two subsections outline methods for estimating the parameters of the prior distribution. The first subsection considers the case when previous test results exist, so data are available to estimate $\hat{\alpha}_1$ and $\hat{\alpha}_2$. The second subsection considers the case when no previous test data are available.

Previous Test Data Exists

When previous test data that are representative of the current test data for a component with perfect test results, this data can be fit to yield parameter estimates $\hat{\alpha}_1$ and $\hat{\alpha}_2$ for the prior beta distribution. Let z_1, z_2, \dots, z_n denote the fraction surviving for previous tests on a component of interest with equal sample sizes (which has perfect test results and need a beta prior distribution). The maximum likelihood estimators satisfy the simultaneous equations (Evans, Hastings, and Peacock, 2000, page 41):

$$\psi(\hat{\alpha}_1) - \psi(\hat{\alpha}_1 + \hat{\alpha}_2) = \frac{1}{n} \sum_{i=1}^n \log z_i,$$

$$\psi(\hat{\alpha}_2) - \psi(\hat{\alpha}_1 + \hat{\alpha}_2) = \frac{1}{n} \sum_{i=1}^n \log(1 - z_i),$$

where ψ is the digamma function. Law and Kelton (2000) outline methods for calculating $\hat{\alpha}_1$ and $\hat{\alpha}_2$. These equations have no closed-form solution and must be solved iteratively. Alternatively, the method-of-moments estimates are found by equating the population mean, μ , and population variance, σ^2 , to the associated sample moments:

$$\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i, \quad s^2 = \frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^2,$$

which results in the closed-form method of moments estimators:

$$\hat{\alpha}_1 = \frac{(1 - \bar{z})\bar{z}^2}{s^2} - \bar{z}, \quad \hat{\alpha}_2 = \frac{\hat{\alpha}_1(1 - \bar{z})}{\bar{z}}.$$

Example: Estimating the Beta Parameters from Previous Experiments

Consider the previous example, where previous test results on component 2 have yielded the following $n = 4$ fractions surviving:

$$z_1 = \frac{27}{28}, \quad z_2 = \frac{28}{28}, \quad z_3 = \frac{26}{28}, \quad z_4 = \frac{27}{28}.$$

Because the sample mean and variance are

$$\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i = \frac{27}{28} \cong 0.964,$$

$$s^2 = \frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^2 = \frac{1}{1568} \cong 0.000638,$$

the method-of-moments estimators are (these correspond to Case I from the previous list):

$$\hat{\alpha}_1 = \frac{1431}{28} \cong 51.11, \quad \hat{\alpha}_2 = \frac{53}{28} \cong 1.89.$$

When these values for the parameters are used in `seriesystembayesboot`, the median of five lower 95% confidence bounds with $B = 10,000$ for the system reliability is 0.763.

No Previous Test Data Exist

We now turn to the more difficult case of determining the prior beta distribution parameter estimates $\hat{\alpha}_1$ and $\hat{\alpha}_2$ in the case of a component with perfect test results and when no previous test data are available. For such a component, the point estimate of the component reliability is $\hat{p} = 1$ and the CP lower reliability bound is $p_L = \alpha^{1/n}$. One heuristic technique for determining the parameters is to choose $\hat{\alpha}_1$ and $\hat{\alpha}_2$ such that $F(p_L) = \alpha$, i.e.,

$$\int_0^{p_L} \frac{\Gamma(\hat{\alpha}_1 + \hat{\alpha}_2)}{\Gamma(\hat{\alpha}_1)\Gamma(\hat{\alpha}_2)} x^{\hat{\alpha}_1 - 1} (1 - x)^{\hat{\alpha}_2 - 1} dx = \alpha. \quad (1)$$

The intuition behind this choice is that 100 α % of the time, a prior component reliability (which will be modified subsequently by the data set) will assume a value less than p_L . One problem with this criteria is that there are an infinite number of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ that satisfy this equation. Further refinement is necessary.

For a sample of size $n > 1$, the $(\hat{\alpha}_1, \hat{\alpha}_2)$ pair satisfying Equation (1) will (a) intersect the line $\hat{\alpha}_1 = 1$ on $0 < \hat{\alpha}_2 < 1$ and (b) intersect the line $\hat{\alpha}_2 = 1$

on $\hat{\alpha}_1 > 1$. One technique for determining a $(\hat{\alpha}_1, \hat{\alpha}_2)$ pair is to find the intersection of the values of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ that satisfy Equation (1) and the lines $\hat{\alpha}_1 = 1$ and $\hat{\alpha}_2 = 1$. These two points of intersection, or any point on the line segment connecting them, can be used as prior beta distribution parameter estimates. It is interesting to note that

- the intersection of Equation (1) and the line $\hat{\alpha}_1 = 1$ corresponds to Case III for the beta distribution parameters (Scenario 1)
- the intersection of Equation (1) and the line $\hat{\alpha}_2 = 1$ corresponds to Case II for the beta distribution parameters (Scenario 2)
- any point on the line segment connecting the two intersection points (not including the endpoints of the segment) corresponds to Case IV for the beta distribution parameters (Scenario 3).

We first consider the intersection of Equation (1) and $\alpha_1 = 1$. Integration of the beta probability density function is analytic in this case, yielding

$$1 - (1 - p_L)^{\hat{\alpha}_2} = \alpha$$

or

$$\hat{\alpha}_2 = \frac{\log(1 - \alpha)}{\log(1 - \alpha^{1/n})}.$$

Next, we first consider the intersection of Equation (1) and $\alpha_2 = 1$. The integration of the beta probability density function is analytic in this case as well, yielding

$$p_L^{\hat{\alpha}_1} = \alpha$$

or

$$\hat{\alpha}_1 = n.$$

Example: Three-Component Series System with Beta Prior Distributions

Consider again the three-component series systems. System 1 has test results displayed in Table 1. System 2 has test results displayed in Table 4. The point estimate for the system reliability of System 1 is

$$\frac{21}{23} \cdot \frac{27}{28} \cdot \frac{82}{84} = \frac{1107}{1288} \cong 0.8595,$$

and the point estimate for the system reliability of System 2 is

$$\frac{21}{23} \cdot \frac{28}{28} \cdot \frac{82}{84} = \frac{41}{46} \cong 0.8913.$$

Hence, the slight difference between the two test results (the perfect test results for component 2) has resulted in a $0.8913 - 0.8595 = 0.0318$ increase in the point estimate for the system reliability. A similar

TABLE 6. Lower Reliability Bounds ($\alpha = 0.05$) for the System Reliability of a Three-Component Series System with Alternative Beta Prior Parameters

Model	$\hat{\alpha}_1$	$\hat{\alpha}_2$	p_L	Δp_L
Uniform prior	1	1	0.748	0.002
Scenario 1	1	0.022418	0.785	0.039
Scenario 2	28	1	0.769	0.023
Scenario 3	14.5	0.511209	0.772	0.026

increase in the lower bound for the system reliability for a reasonable procedure is expected.

The earlier analysis of System 1 using APPL has resulted in an exact (no resampling variability) bootstrap 95% lower limit on the system reliability of 0.746. Table 6 contains 95% lower confidence limits for the system reliability using four different combinations of prior beta distribution parameter estimates $\hat{\alpha}_1$ and $\hat{\alpha}_2$ for component 2. The parameter estimates for Scenario 3 are found by averaging the parameter estimates for Scenarios 1 and 2. The lower bounds p_L are determined by taking the median result of five runs with $B = 10,000$ replications using the bootstrap procedure described earlier. The column labeled Δp_L gives the difference between the lower confidence limit for System 2 and the lower confidence limit for System 1. The uniform prior model is too wide because the $0.748 - 0.746 = 0.002$ increase in the lower bound is inconsistent with the 0.0318 increase in the point estimate for the system reliability. Based on this example only, Scenarios 1 and 3 seem to be the most appropriate because their increases in the lower bound bracket the increase in the point estimator for the system reliability.

Our heuristic, which chooses $\hat{\alpha}_1$ and $\hat{\alpha}_2$ such that $F(p_L) = \alpha$ works reasonably well in the example with one component having perfect test results, but will likely need to be modified if several components have perfect test results. A large-scale Monte Carlo simulation, which involves varying α , the number of system components, the configuration of the system components, and the expected fraction of cases where perfect test results are encountered, is the only way to evaluate the techniques presented here and to compare them, for example, with the asymptotic techniques presented in Mann, Schaefer, and Singpurwalla (1974, p. 498). Such a simulation is appropriate on a system-by-system basis.

Simulation

Monte Carlo simulation is used to test several heuristic methods along with the techniques developed in this paper. We begin with a pilot simulation that is used to evaluate a large number of methods in order to thin the number of methods considered.

Simulation Study A

The system considered in this pilot study is a three-component series system with identical components. In keeping with the earlier example, there are $n_1 = 23$, $n_2 = 28$, and $n_3 = 84$ components of each type placed on test. There are $B = 1000$ bootstrap replications used and 1000 simulation replications conducted. The stated coverage of the lower confidence-interval bound for the system reliability is 0.95. If the intervals cover approximately 95% of the true system reliability values for a wide range of true component reliabilities, then the confidence-interval procedure is performing adequately. Using the two-sided CP confidence-interval procedure for a single component described earlier, the acceptable range for the fraction of simulated confidence intervals (at $\alpha = 0.01$) covering the true system reliability is from 0.931 to 0.968 inclusive. The simulations are run in S-Plus using the `set.seed` command prior to each run to exploit the common random numbers variance reduction technique.

In addition, the number of components that achieve perfect test results is computed for general true component reliabilities p_1 , p_2 , and p_3 . For general n_1 , n_2 , and n_3 , let the random variable W be the number of components with perfect test results. The probability density function of W is

$$f(w) = \begin{cases} (1 - p_1^{n_1})(1 - p_2^{n_2})(1 - p_3^{n_3}) & w = 0 \\ (1 - p_1^{n_1})(1 - p_2^{n_2})p_3^{n_3} \\ \quad + (1 - p_1^{n_1})p_2^{n_2}(1 - p_3^{n_3}) \\ \quad + p_1^{n_1}(1 - p_2^{n_2})(1 - p_3^{n_3}) & w = 1 \\ (1 - p_1^{n_1})p_2^{n_2}p_3^{n_3} \\ \quad + p_1^{n_1}(1 - p_2^{n_2})p_3^{n_3} \\ \quad + p_1^{n_1}p_2^{n_2}(1 - p_3^{n_3}) & w = 2 \\ p_1^{n_1}p_2^{n_2}p_3^{n_3} & w = 3. \end{cases}$$

These values are computed and given in Table 7 for various true, identical component reliabilities ranging from 0.60 to 0.99.

Nine algorithms for handling the case of one or more components with perfect test results are compared in the pilot simulation. We have included algorithms of an ad hoc nature (e.g., Algorithms 2 and

TABLE 7. Estimated Lower Confidence Interval Coverage ($\alpha = 0.05$) for the System Reliability of a Three-Component Series System with 1000 Replications Using Bootstrapping with a Beta Prior Distribution for Perfect Test Results. The Random Variable W Denotes the Number of Components with Perfect Test Results. The Tabled Values Give the Fraction of Intervals that Fall Below the True System Reliability. Fractions Set in Boldface Type Are in the Range 0.931 to 0.968 Inclusive and Are Not Statistically Different from the Stated Coverage of 0.95

True reliability:	0.60	0.70	0.80	0.90	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99
Pr($W = 0$)	0.999	0.999	0.992	0.864	0.822	0.767	0.704	0.621	0.521	0.401	0.267	0.131	0.029
Pr($W = 1$)	10^{-5}	10^{-4}	0.008	0.132	0.170	0.216	0.271	0.334	0.401	0.460	0.483	0.424	0.221
Pr($W = 2$)	10^{-12}	10^{-8}	10^{-5}	0.005	0.008	0.014	0.025	0.044	0.078	0.136	0.234	0.380	0.492
Pr($W = 3$)	10^{-30}	10^{-21}	10^{-13}	10^{-6}	10^{-6}	10^{-5}	10^{-4}	10^{-4}	0.001	0.004	0.016	0.065	0.257
Algorithm 1	0.958	0.964	0.946	0.912	0.910	0.910	0.899	0.899	0.878	0.873	0.809	0.765	0.738
Algorithm 2	0.958	0.964	0.945	0.950	0.960	0.980	0.988	1.000	1.000	1.000	1.000	1.000	1.000
Algorithm 3	0.958	0.964	0.953	0.939	0.954	0.937	0.955	0.972	0.981	1.000	1.000	1.000	1.000
Algorithm 4	0.958	0.964	0.948	0.960	0.957	0.968	0.988	1.000	1.000	1.000	1.000	1.000	1.000
Algorithm 5	0.958	0.964	0.939	0.927	0.918	0.919	0.902	0.910	0.895	0.877	0.832	0.820	0.745
Algorithm 6	0.958	0.964	0.947	0.923	0.942	0.951	0.947	0.947	0.970	0.976	1.000	1.000	1.000
Algorithm 7	0.958	0.964	0.933	0.923	0.947	0.917	0.944	0.942	0.936	0.936	0.982	1.000	1.000
Algorithm 8	0.958	0.964	0.951	0.925	0.925	0.916	0.911	0.926	0.900	0.906	0.876	0.827	1.000
Algorithm 9	0.975	0.947	0.949	0.920	0.918	0.919	0.888	0.897	0.886	0.867	0.822	0.737	0.730

3) and those with some theoretical basis (e.g., Algorithm 9) in order to show that the beta prior approach dominates the other approaches as component reliability increases.

- **Algorithm 1:** Pure bootstrapping. A component with a perfect test always generates perfect simulated results.
- **Algorithm 2:** Always assume a failure. When component i has perfect test results (i.e., $y_i = n_i$), introduce an artificial failure by assuming that $y_i = n_i - 1$, for $i = 1, 2, \dots, n$.
- **Algorithm 3:** Increase the sample size. For perfect test results, artificially increase sample size to approximate the lower confidence bounds with a single failure using the confidence intervals for a single component given earlier in the paper, then bootstrap. In our case, $n_1 = 37$ ($y_1 = 36$), $n_2 = 45$ ($y_2 = 44$), and $n_3 = 134$ ($y_3 = 133$).

- **Algorithm 4:** Bayes bootstrapping with $\alpha_1 = 1$ and $\alpha_2 = 1$ (i.e., uniform prior).
- **Algorithm 5:** Bayes bootstrapping with $\alpha_1 = 1$ for all components, and $\alpha_2 = \log(1 - \alpha) / \log(1 - \alpha^{1/n_i})$, for $i = 1, 2, \dots, k$, as described earlier.
- **Algorithm 6:** Bayes bootstrapping with $\alpha_1 = n_i$, for $i = 1, 2, \dots, k$, and $\alpha_2 = 1$ for all components, as described earlier.
- **Algorithm 7:** Bayes bootstrapping with α_1 and α_2 that are averages of the values given in Algorithms 5 and 6, as described earlier.
- **Algorithm 8:** Bayes bootstrapping with $\alpha_1 = 100$ and $\alpha_2 = 1$.
- **Algorithm 9:** A procedure from Mann, Schafer, and Singpurwalla (1974, pp. 497–499) which, using asymptotic normal theory, calculates a lower

bound as

$$\prod_{i=1}^k \frac{y_i}{n_i} - z_\alpha \sqrt{\prod_{i=1}^k \left(\frac{y_i}{n_i}\right)^2 \sum_{j=1}^k \left(\frac{1}{y_j} - \frac{1}{n_j}\right)}.$$

The performance of the confidence intervals given in Table 7 is as expected. Algorithm 1, for example, which takes the overly optimistic, pure bootstrapping approach, produces lower confidence limits that are shifted up, resulting in fewer than expected lower confidence limits that fall below the true system reliability. The opposite case is true for the pessimistic uniform prior distribution in Algorithm 4. In fact, once the Bayesian portion of the algorithm began to dominate (i.e., when the component reliabilities are large), all nine algorithms fail to deliver confidence intervals with the appropriate coverage. We experimented with the confidence interval that performed the best (Algorithm 7, which averages the parameter estimates of Algorithms 5 and 6) by replacing average with linear combination, but did not produce results that were significantly better than those presented in Table 7.

The abysmal performance of all of these algorithms for high-reliability components is consistent with the work of Martz and Duran (1985), who considered lower confidence bounds for the system reliability of 20 system configurations and component reliabilities using three algorithms and two values of α (0.05 and 0.10). Their intervals also diverged from the stated coverages.

Simulation Study B

This poor performance led us to recode our algorithms in C and to do an exhaustive search in the (α_1, α_2) plane for values of the beta prior parameters α_1 and α_2 that yield reasonable coverages for lower confidence bounds on the system reliability. We returned to the case of a single component. Figure 4 shows the results of this exhaustive search for $n = 23$ components on test. Every (α_1, α_2) pair that resulted in a confidence interval whose coverage did not statistically differ from 0.95 was plotted for $p = 0.91, 0.92, \dots, 0.99$. For each particular population reliability p shown in Figure 4, the areas where appropriate coverages are achieved are quite narrow. Unfortunately, the graph in Figure 4 shows that there is no single (α_1, α_2) pair that will work for all values of p .

The following procedure has been developed as a compromise that allows reasonable lower confidence-

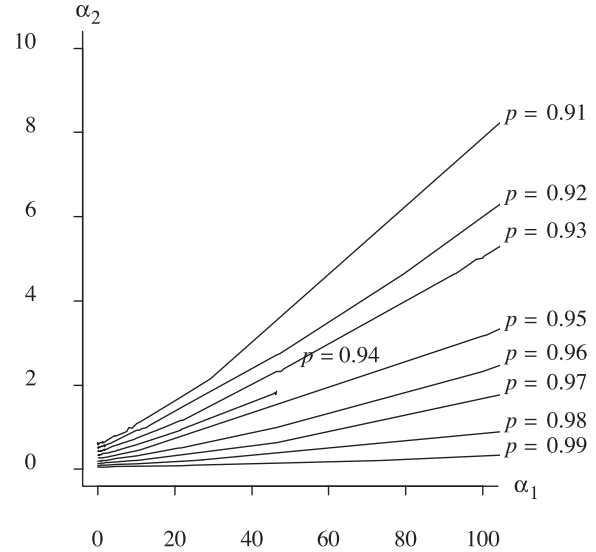


FIGURE 4. Prior Distribution Parameter Pairs that Give Accurate Coverage for the Example as a Function of the Reliability p When $n = 23$.

limit coverage in the case of a system with one or more components having perfect test results:

- For each of the components in the system, consult with someone familiar with the component to get a point estimate of the component reliability p_i^* , $i = 1, 2, \dots, k$.
- Determine the number of components to be tested n_1, n_2, \dots, n_k .
- For each (p_i^*, n_i) pair with perfect test results, perform an exhaustive search of the (α_1, α_2) plane to find a $(\tilde{\alpha}_1, \tilde{\alpha}_2)$ pair that yields appropriate coverage.
- Perform Bayesian bootstrapping as described earlier in this paper using the test results (n_i, y_i) and the appropriate $(\tilde{\alpha}_1, \tilde{\alpha}_2)$ values from the previous step.

The final example illustrates this technique for a single-component system and three-component system.

Example: Single-Component System and Three-Component Series System

Figure 5 shows the actual coverage for 1000 simulation replications for a single-component system with $n = 28$ and a three-component system with $n_1 = 23, n_2 = 28,$ and $n_3 = 84$. All true component

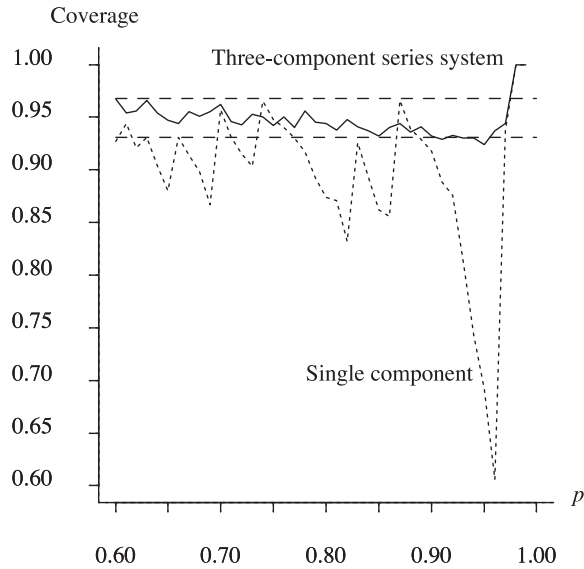


FIGURE 5. Lower 95% Confidence Bound Coverage for a Single Component System (dotted) and a Three-Component System (solid) and Region Not Statistically Different from the Specification (dashed).

reliabilities are equal and are plotted on the horizontal axis. The stated coverage on all intervals is 0.95. The usual bounds around 0.95 (at 0.931 and 0.968), which denote confidence intervals of which actual coverage does not differ significantly from the specification are given as horizontal dashed lines. All Bayesian procedures use $(\alpha_1, \alpha_2) = (252.28, 4.67)$, which were values that fell outside of the axes in Figure 4 associated with the $p^* = 0.97$ estimate for the reliability of the second component. The jagged appearance for the coverage for the interval for a single component (dotted) is consistent with the same pattern shown by Blyth (1986). The three-component system (solid), on the other hand, has different numbers of components on test that seem to average out these fluctuations, resulting in appropriate coverage through $p = 0.97$.

Conclusion

Determining lower confidence bounds from binary data remains an important yet elusive task. The Bayesian bootstrapping procedures developed here yield adequate coverages given that an expert is able make a good initial estimate of the reliabilities of individual components. The estimates discussed here improve with increasing system complexity.

Acknowledgments

The author gratefully acknowledges the helpful interactions with John Backes, Don Campbell, Mike Crawford, Mark Eaton, Jerry Ellis, Greg Gruver, Jim Scott, Bob Shumaker, Bill Treadwell, and Mark Vinson concerning the paper and helpful comments from two anonymous referees and the editor. The author also thanks Diane Evans and Andy Glen for their help with the development of APPL.

Appendix 1

S-Plus code for calculating a CP $(1-\alpha)100\%$ lower confidence bound for a single component for n components on test and y passes.

```
confintlower <- function(n, y, alpha) {
  if (y == 0) {
    pl <- 0
  }
  if (y == n) {
    pl <- alpha ^ (1 / n)
  }
  if (y > 0 && y < n) {
    fcrit1 <- qf(alpha, 2 * y, 2 * (n - y + 1))
    pl <- 1 / (1 + (n - y + 1) / (y * fcrit1))
  }
  pl
}
```

Appendix 2

S-Plus code for calculating a bootstrap $(1-\alpha)100\%$ lower confidence interval bound for a k -component series system of independent components using B bootstrap replications.

```
seriessystemboot <- function(n, y, alpha) {
  k <- length(n)
  b <- 10000
  z <- rep(1, b)

  point <- prod(y) / prod(n)

  for (j in 1:b) {
    for (i in 1:k) {
      z[j] <- z[j] * rbinom(1, n[i], y[i] / n[i])
        / n[i]
    }
  }
  z <- sort(z)
  pl <- z[floor(alpha * b)]
  c(point, pl)
}
```

Appendix 3

APPL code for calculating a bootstrap $(1-\alpha)100\%$ lower confidence interval bound for a k -component series system of independent components

using the equivalent of $B = +\infty$ bootstrap replications.

```
n1 := 23;
y1 := 21;
X1 := BinomialRV(n1, y1 / n1);
X1 := Transform(X1, [[x -> x / n1],
  [-infinity, infinity]]);
n2 := 28;
y2 := 27;
X2 := BinomialRV(n2, y2 / n2);
X2 := Transform(X2, [[x -> x / n2],
  [-infinity, infinity]]);
n3 := 84;
y3 := 82;
X3 := BinomialRV(n3, y3 / n3);
X3 := Transform(X3, [[x -> x / n3],
  [-infinity, infinity]]);
Temp := Product(X1, X2);
T := Product(Temp, X3);
```

Appendix 4

S-Plus code for calculating a bootstrap $(1 - \alpha)100\%$ lower confidence interval bound for a k -component series system of independent components with some perfect component test results using B bootstrap replications.

```
seriessystembayesboot <- function(n, y, alpha) {
  k <- length(n)
  alpha1 <- 1
  alpha2 <- 1
  b <- 10000
  z <- rep(1, b)

  point <- prod(y) / prod(n)

  for (j in 1:b) {
    for (i in 1:k) {
      if (y[i] == n[i]) z[j] <- z[j]
        * rbeta(1, alpha1 + y[i], alpha2)
      else z[j] <- z[j]
        * rbinom(1, n[i], y[i] / n[i]) / n[i]
    }
  }
}
```

```
z <- sort(z)
p1 <- z[floor(alpha * b)]
c(point, p1)
}
```

References

- AGRESTI, A. and COULL, B. (1998). "Approximate is Better than 'Exact' for Interval Estimation of Binomial Proportions". *The American Statistician* 52, pp. 119–126.
- BLYTH, C. R. (1986). "Approximate Binomial Confidence Limits". *Journal of the American Statistical Association* 81, pp. 843–855; Corrigenda (1989), 84, p. 636.
- CHICK, S. (2001). Personal communication.
- EFRON, B. and TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York, NY.
- EVANS, M.; HASTINGS, N.; and PEACOCK, B. (2000). *Statistical Distributions, 3rd ed.* John Wiley & Sons, New York, NY.
- GLEN, A. G.; EVANS, D. E.; and LEEMIS, L. M. (2001). "APPL: A Probability Programming Language". *The American Statistician* 55, pp. 156–166.
- LAW, A. M. and KELTON, W. D. (2000). *Simulation Modeling and Analysis, 3rd ed.* McGraw-Hill, New York, NY.
- LEEMIS, L. M. and TRIVEDI, K. S. (1996). "A Comparison of Approximate Interval Estimators for the Bernoulli Parameter". *The American Statistician* 50, pp. 63–68.
- MANN, N. R.; SCHAFER, R. E.; and SINGPURWALLA, N. D. (1974). *Methods for Statistical Analysis of Reliability and Life Data*. John Wiley & Sons, New York, NY.
- MARTIN, M. A. (1990). "On Bootstrap Iteration for Coverage Correction in Confidence Intervals". *Journal of the American Statistical Association* 85, pp. 1105–1118.
- MARTZ, H. F. and DURAN, B. S. (1985). "A Comparison of Three Methods for Calculating Lower Confidence Limits on System Reliability Using Binomial Component Data". *IEEE Transactions on Reliability* R-34, pp. 113–120.
- MARTZ, H. F. and WALLER, R. A. (1982). *Bayesian Reliability Analysis*. John Wiley & Sons, New York, NY.
- NEWCOMBE, R. G. (2001). "Logit Confidence Intervals and the Inverse Sinh Transformation". *The American Statistician* 55, pp. 200–202.
- PADGETT, W. J. and TOMLINSON, M. A. (2003). "Lower Confidence Bounds for Percentiles of Weibull and Birnbaum-Saunders Distributions". *Journal of Statistical Computation and Simulation* 73, pp. 429–443.
- VOLLSET, S. E. (1993). "Confidence Intervals for a Binomial Proportion". *Statistics in Medicine* 12, pp. 809–824.

