

NIST Technical Note 2119

**Estimating Instrument Performance:
with Confidence Intervals and
Confidence Bounds**

Zhan-Qian John Lu

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.TN.2119>

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

NIST Technical Note 2119

Estimating Instrument Performance: with Confidence Intervals and Confidence Bounds

Zhan-Qian John Lu
*Division of Statistical Engineering
Information Technology Laboratory*

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.TN.2119>

September 2020



U.S. Department of Commerce
Wilbur L. Ross, Jr., Secretary

National Institute of Standards and Technology
Walter Copan, NIST Director and Undersecretary of Commerce for Standards and Technology

Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

National Institute of Standards and Technology Technical Note 2119
Natl. Inst. Stand. Technol. Tech. Note 2119, 54 pages (September 2020)
CODEN: NTNOEF

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.TN.2119>

Abstract

Instrument performance is evaluated through performing a set of tests, and the test results determine whether the instrument meets or fails the standard requirements. The performance test data are analyzed and usually point estimate and confidence bounds are calculated, on which a testing decision is made. This chapter provides recipes of statistical analysis of the test data, including estimation of probability of detection or identification, or false alarm rate, and particularly on how various confidence intervals and confidence bounds are computed based on sound statistical principles.

Key words

Probability of detection; binomial distribution; confidence intervals; confidence bounds; probability distribution estimation; Bayesian credible intervals.

Table of Contents

1	Introduction	1
1.1	Estimating Probability of Detection or Identification	1
1.2	Confidence Intervals and Choosing Among Them	2
1.3	Statistical Treatment of Quantitative Responses	3
1.4	Use of Confidence Intervals in Decision Making	4
2	Confidence Intervals When Sample Size Is Large For Binary Responses	5
2.1	The Wald Interval and Bounds	5
2.2	Sample Size Design	6
2.3	The Wilson Interval and Bounds	7
2.4	Agresti-Coull Interval and Bounds	8
2.5	Poisson Adjustment For Rare Events	8
3	Probabilistic Bases and Interpretations of Interval Estimation	10
3.1	Binomial Model	10
3.2	Normal Approximation	11
3.3	Poisson Approximation	12
3.4	Asymptotic Coverage and Relation to Testing	13
4	Exact Interval and Bayesian Intervals	15
4.1	Clopper-Pearson Interval	15
4.2	Bayesian Inference	16
4.3	Bayesian Posterior Summaries	17
4.4	Prior Choice and Incorporating External Information	18
4.5	Bayesian Credible Intervals	21
4.6	Comparison Among the Exact and Bayesian Intervals	23
5	Statistical Analysis of Quantitative Data	28
5.1	Confidence Intervals for Parameters	28
5.2	Parametric Approach to Probability Estimation	30
5.3	Sample Quantiles and Related Confidence Intervals	32
	Appendix A: Technical Notes and Calculation Details	34
	Appendix B: Bayesian Intervals Computation and Tables	35
	References	36

List of Figures

Figure 1	Display of Binomial Probabilities and Normal Approximations.	11
Figure 2	Display of Three Beta Distributions.	20
Figure 3	Illustration of Bayesian Inference: Data $n = 20, s = 2$ and Uses of Agresti and Jeffreys priors Are Compared to the Likelihood.	22
Figure 4	Illustration of Bayesian Inference: Data $n = 20, s = 17$ and Uses of Agresti and Jeffreys priors Are Compared to the Likelihood.	23
Figure 5	Comparison of Interval Estimates Using Clopper-Pearson Exact Method, Jeffreys Credible Intervals and Bayesian Likelihood Intervals.	25
Figure 6	Comparison of Bayesian Credible Intervals: with Jeffreys, Uniform, and Agresti Priors	26
Figure 7	Display of Empirical CDF of Test Data in Example 10 and the Normal Probability Model-based Fit.	31

List of Tables

Table 1	Tables of Select Quantiles of χ^2 Distribution For $\alpha = 0.1$	9
Table 2	Probability table of CDF $Bin(20, 0.95)$	10
Table 3	Approximation of CDF $Bin(20, 0.1)$ by $N(\mu - 0.5, \sigma^2)$ and $Pois(\lambda)$, Where $\mu = \lambda = 20 \times 0.1 = 2, \sigma^2 = 20 \times .1 \times .9 = 1.8$	13
Table 4	Bayesian Inference of Binomial Data Under Select Prior Choices	19
Table 5	95% Bayesian Credible Intervals $[LL, UL]$ (with median value Med) Under Three Prior Choices for Observed Data $n = 20, s = 2$ (columns 2,3,4) and $n = 20, s = 17$ (columns 5,6,7)	24
Table 6	95% Limits $[LL, UL]$ for Bayesian Intervals for Binomial Data with Sample Size $n = 7$ Through 18	38
Table 7	95% Limits $[LL, UL]$ for Bayesian Intervals for Binomial Data with Sample Size $n = 19$ Through 24	39
Table 8	95% Limits $[LL, UL]$ for Bayesian Intervals for Binomial Data with Sample Size $n = 25$ Through 30	40
Table 9	90% Limits $[LL, UL]$ for Bayesian Intervals with Sample Size $n=7$ Through 18.	41
Table 10	90% Limits $[LL, UL]$ for Bayesian Intervals with Sample Size $n=19$ Through 24.	42
Table 11	90% Limits $[LL, UL]$ for Bayesian Intervals with Sample Size $n=25$ Through 30.	43
Table 12	80% Equal-tailed Bayesian Intervals For $n = 7$ Through 18	44
Table 13	80% Equal-tailed Bayesian Intervals For $n = 19$ Through 24	45
Table 14	80% Equal-tailed Bayesian Intervals For $n = 25$ Through 30	46

Example

List of Examples

1	Wald Intervals.	5
2	Wald Confidence Bounds	6
3	Sample Size Requirements.	6
4	Wilson Intervals and Bounds.	7
5	Agresti-Coull intervals.	8
6	Poisson Adjustment For Rare Events.	9
7	Relating Confidence Bound to Testing Decision.	14
8	Clopper-Pearson Exact Interval.	16
9	Bayesian and Likelihood Intervals.	24
10	Confidence Intervals For Parameters.	29
11	Parametric Probability Estimation.	30
12	Confidence Intervals For Quantiles	34

1. Introduction

Performance of an instrument or device is evaluated through a suite of tests. Decision on whether an instrument meets the requirements depends on satisfying a set of performance metrics or scoring criteria. The performance metrics typically include the **probability of identification**, i.e. the chance when an underlying source such as a particular type of radionuclide is correctly identified by a detector; the **probability of detection**, the chance that the presence of known source material is detected under specified distance (closeness to the source) and time limits (such as 15 seconds in real-time) by a detecting instrument; and the **probability of false positive**, the probability of giving a positive response when the underlying source is not present. The decision scoring criteria include point estimation of the performance metrics and whether the associated confidence bounds satisfy the pre-specified requirements. This chapter provides recipes for statistical analyses of the performance test data, including estimation of probability of detection or identification, and how various confidence intervals and confidence bounds are computed based on binary measurement data. Measurements of continuous and other performance measures have also been strongly recommended due to the potential to employ more advanced statistical tools including parametric models and the related statistical analysis is also reviewed.

1.1 Estimating Probability of Detection or Identification

In this section, we use the generic notation p , to denote the probability of identification P_{ID} , or the probability of detection, P_D . Estimation of p and confidence bounds or intervals for p are briefly described when using performance test data. The estimation of false alarm rate is treated similarly, and the difference is that while it is desirable to have an instrument or testing procedure with a P_{ID} or P_D close to 1, the sacrifice is to incur an intolerable false alarm rate P_{FA} , which should be minimized to an acceptable level close to 0 at the same time.

If the test data from a given instrument performance evaluation is derived from n independent binary trials, then denote by S or S_n , the total number of positive identifications or detections from n tests. That is, $S_n = \sum_{i=1}^n x_i$ where x_1, \dots, x_n represent test results, each x_i taking on value 0 for failure and 1 for success. It should be noted that we also use s denote the observed or realized value of S , which is reserved for treating it as a random variable. The point estimate of p is given by

$$\hat{p} = S/n, \quad (1)$$

which is simply an alternative expression of the familiar sample mean formula $\hat{p} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, where n is the total number of tests.

That \hat{p} is a reasonable estimate of p is based on the fact that it is an unbiased estimator of p , that is, its expected value is exactly p . As an unbiased estimator, its variability or uncertainty is usually characterized through the standard deviation,

$$sd(\hat{p}) = \sqrt{p(1-p)/n}. \quad (2)$$

An estimate of (2), called sample standard deviation, is given by substituting p with \hat{p} .¹

The availability of sample standard uncertainty does not guarantee making a probabilistic statement on the likely values of the estimator \hat{p} . More specifically, in order to make a statement that some interval to be constructed, based on the performance test data, that the interval will cover the true p will high probability, say $1 - \alpha$, for some small $0 < \alpha < 1$, one needs additional information on the sampling distribution of \hat{p} for any given performance data.

1.2 Confidence Intervals and Choosing Among Them

The estimation of binomial probability p is based on S , which has the binomial distribution, and so the sampling distribution of \hat{p} is very close to the normal distribution with mean p and standard deviation as given in (2) when the sample size n is large (Section 3). Based on the latter property of asymptotic normality, one can easily derive at least three confidence intervals, including Wald, Wilson, and Agresti-Coull intervals as discussed in Section 2. These intervals, however, all depend on the goodness of the normal approximation to the binomial distribution (see Section 3.2 of this Chapter) and so they belong to large sample approximate methods. For larger n , say bigger than 40, there is not much difference among them as Wilson, Jeffreys and the Agresti-Coull intervals are all comparable, and the Agresti-Coull interval is the simplest to present (see for example [2]. Though the standard Wald interval is often discussed in elementary textbooks, caution should be exercised as it may not be appropriate to use in many situations even when n is fairly large, as pointed out by several studies, as reviewed in [6]. When $n \leq 40$, there is some significant difference among the large sample methods, and the Wilson or Agresti-Coull interval may be preferred over the standard Wald interval. Especially when the observed data very small or very close to n , the Poisson adjustment procedure of Subsection 2.5 is also recommended. Furthermore, for small sample size n , say less than 20 or 40, we recommend the exact and Bayesian methods which will be discussed in Section 4, for which recommendation on the Bayesian credible intervals is given in Section 4.6. Choosing a prior distribution on p represents both a challenge and also an added flexibility with the Bayesian approach, as for example, with one of three noninformative prior choices, such as Jeffreys prior, uniform (or $\text{Be}(0.5, 0.5)$), and Agresti prior (Table 4), the corresponding Bayesian credible intervals approximate most of the classical confidence intervals in Section 2, and so the Bayesian approach may be considered as a more powerful and general approach for binomial probability inference.

¹Strictly speaking, the sample version of (2), has the alternative expression

$$\sqrt{\hat{p}(1 - \hat{p})/n} = \sqrt{n^{-2} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{n-1}{n^2}} D$$

where $D = \sqrt{(n-1)^{-1} \sum_{i=1}^n (x_i - \bar{x})^2}$. So, by differing a factor $\sqrt{\frac{n-1}{n}}$, it is not quite the same, but very close to $n^{-1/2}D$, the sample standard deviation for \bar{x} , as discussed in Subsection 1.3 for non-binary data.

1.3 Statistical Treatment of Quantitative Responses

Quantitative responses (actual performances) are sometimes recorded in testing, such as the response time to detection or identification. Sometimes, the performance metrics are not categorical decisions, such as the closeness or distance to target. The recording of the actual performance of an instrument can lead to several advantages, first to characterize the instrument's actual performance (not just simply pass or fail decision), and also to have data which can be combined or compared with experiments in another time or with a different instrument. Indeed, the 1998 National Research Council (NRC) report [13] recommended this continuous paradigm in order to better take advantage of modern statistical methods. Parametric model approach can provide alternative way of probability estimation with better precision even with a smaller sample size, if a parametric model can be found to fit the continuous data well and is deemed adequate.

When the test data, say x_1, x_2, \dots, x_n are non-binary responses from n tests, one may be interested in the mean μ and standard deviation σ or even overall shape of the distribution for the data. Sample mean is given by

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

The sample standard deviation is given by

$$D = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Because $\text{var}(\bar{x}) = \frac{1}{n} \sigma^2$ when the data x_1, x_2, \dots, x_n are independent and have the same distribution with mean μ and variance σ^2 , an estimate for $\text{var}(\bar{x})$ is given by: $\frac{1}{n} D^2$.

One may be interested in the estimation of cumulative distribution function (CDF), $F(x) = P(X \leq x)$, where X represents the population from which the test data come from. A parametric model approach is given by

$$\hat{F}(x) = G_{\bar{x}, D}(x) \tag{3}$$

where $G_{\mu, \sigma}$ is the CDF of some standard probability models. For example, $G_{\mu, \sigma} = \Phi\left(\frac{x-\mu}{\sigma}\right)$ if X follows a normal distribution $N(\mu, \sigma^2)$, and $G_{\mu, \sigma} = \Phi\left(\frac{\log(x)-\mu}{\sigma}\right)$ if X follows the log-normal distribution, $\log(X) \sim N(\mu, \sigma^2)$. Here Φ is the CDF of standard normal distribution as defined in (14). This topic will be followed up in Subsection 5.2, and in the rest of Section 5, confidence intervals for both parametric methods and nonparametric approaches using quantiles and order statistics will be discussed as well.

We should point out that even though this chapter's focus is on characterizing the performance of a single instrument, there are clear parallel approaches for comparing two or more instruments, using the related confidence intervals or testing procedures, see [17] where comparing instruments for both binary and continuous data are given.

1.4 Use of Confidence Intervals in Decision Making

In practice, even when the main interest is to test hypotheses about a parameter μ such as when a probability $\mu(=p) = \mu_0$ for some pre-specified threshold μ_0 is of interest, confidence intervals provide not only a way of summarizing data evaluation but also a tool for decision making as well. Given data vector $\mathbf{X} = (x_1, \dots, x_n)$ and a confidence interval $C(X)$ based on it, one can test the null hypothesis

$$H_0 : \mu = \mu_0 \text{ versus } H_1 : \mu \neq \mu_0$$

for every μ_0 without any recalculation, by rejecting H_0 if and only if $\mu_0 \notin C(\mathbf{X})$, that is μ_0 does not belong to $C(\mathbf{X})$. Furthermore, given a $(1 - \alpha)100\%$ confidence interval for μ , the test based on the confidence interval has significance level at most α .

The reason for reporting confidence intervals, perhaps more important in practice, is that a confidence interval can differentiate between statistical significance and practical significance. One can construct examples when μ is practically the same as μ_0 and yet H_0 is rejected, simply because the sample size is large. Confidence intervals are proposed as a representative approach for testing equivalence [11]. The equivalence test is widely used in pharmaceutical problems and the confidence interval representation to equivalence testing is useful in comparing a new instrument against a standard or reference instrument or comparing multiple instruments [17].

In constructing confidence intervals, one not only requires that the stated nominal confidence level is as close to the true coverage as possible, one would also like the confidence interval to be as precise and as short as possible. The essential decision problem in product performance testing is the following: deciding on the careful balance between accepting a product which meets the performance target versus rejecting an inferior product with adequate frequency. The way around this issue is by finding a decision rule using either a lower confidence interval, say 80% confidence level for compliance, which will lead to shorter interval so a more stringent criterion of acceptance. Also increase of sample size is another way of improving the testing performance as the confidence intervals will be more precise and less vague, and thus a more powerful test. The precision of confidence interval is closely related to evaluation of the power of test, as discussed in [10].

2. Confidence Intervals When Sample Size Is Large For Binary Responses

In this section, we discuss several commonly used confidence intervals or bounds for probability estimation based on binary performance data. They are all fairly easy to compute and use, and have been recommended or used in different contexts in applications.

2.1 The Wald Interval and Bounds

The standard Wald confidence interval for p is defined as

$$\hat{p} \pm z_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}. \quad (4)$$

where z_{1-c} denotes the coverage factor, given by the $100(1-c)\%$ upper quantile of the standard normal distribution. Some commonly used values for z include

$$z_{.90} = 1.281552, z_{.95} = 1.644854, z_{.975} = 1.959964$$

which correspond to the 80%, 90% and 95% confidence intervals, respectively.

Sometimes, the *continuity-corrected* Wald interval is used, given by:

$$\hat{p} - z_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n} - \frac{1}{2n} \leq p \leq \hat{p} + z_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n} + \frac{1}{2n}. \quad (5)$$

However, the interval given by (5) is wider and thus more conservative than (4).

Example 1 Wald Intervals.

Suppose in a detection experiments we have $n = 20, s = 17$. $\hat{p} = 18/20 = 0.85$. The sample standard deviation for \hat{p} is $\sqrt{0.85 \times (1 - 0.85)/20} = 0.079$. The standard 90% Wald confidence interval is given by $[0.85 \pm 1.645 \times .079] \approx [0.72, 0.98]$. Adding the continuity correction factor $1/(2 \times 20) = 0.025$ to the lower endpoint and subtracting it to the upper endpoint, we obtain the corrected Wald interval: $[\cdot72 - .025, .98 + 0.025] \approx [0.695, 1.005]$.

Both (4) and (5) give the $100(1-\alpha)\%$ confidence interval for p and (5) is a significant improvement over (4) when sample size n is moderate, but the continuity corrected confidence interval is wider and *more conservative*. The Wald interval theory is based on estimating the standard uncertainty (2) by substituting \hat{p} for p and approximate normality of \hat{p} when n is large in Section 3.2. The Wald interval (4) is widely used because it is easy to use: intervals with a different confidence level $1-\alpha$ simply involves multiplying a *different coverage factor*, i.e. $z_{1-\alpha/2}$, with the same estimated standard uncertainty, an approach widely adopted in standard metrology practice such as ISO GUM.

Wald confidence bounds: The Wald $100(1-\alpha)\%$ **upper confidence bound** with *continuity correction* is given by

$$p : p \leq \hat{p} + z_{1-\alpha} \sqrt{\hat{p}(1-\hat{p})/n} + 1/(2n).$$

The $100(1 - \alpha)\%$ **lower confidence bound** with continuity correction is

$$p : p \geq \hat{p} - z_{1-\alpha} \sqrt{\hat{p}(1 - \hat{p})/n} - 1/(2n).$$

The standard Wald confidence bounds are defined without the continuity correction factor $1/(2n)$. Note that the critical factor $z_{1-\alpha}$ is about 1.645 for $\alpha = 0.05$ and 1.28 for $\alpha = 0.10$. The lower and upper bounds form the bounds for the two-sided $100(1 - 2\alpha)\%$ Wald confidence interval.

Example 2 Wald Confidence Bounds

Suppose in a detection experiment, we obtain $n = 20, s = 17$. We already have $\hat{p} = 0.85$, and sample deviation for \hat{p} is 0.079 from Example 1. Since $z_{0.90} = 1.28$, we calculate the 90% lower bound: $p > 0.85 - 1.28 \times 0.079 \approx 0.749$, and with continuity correction of $1/(2 \times 20) = 0.025$, the lower bound is $0.749 - 0.025 = 0.724$. The 90% upper bound is: $p < 0.85 + 1.28 \times 0.079 \approx 0.951$, and with continuity correction $p < 0.85 + 1.28 \times 0.079 + 0.025 \approx 0.976$.

2.2 Sample Size Design

The requirements for confidence intervals are two-folds: not only that they should have good probability coverage over many different experiments and applications, but also they should be as accurate as possible. That is, too wide an interval estimate may be useless. One way to achieve an interval estimate to acceptable level of accuracy is to increase the sample size at the design of experiments stage (prior to data collection). One can also do sequential experiments but they will not be discussed here. Discussion of sample size impacts other confidence intervals as well but for simplicity we just focus on the Wald intervals, and the sample size consideration impacts other interval estimates comparably.

Example 3 Sample Size Requirements.

If we know that $p^* = 0.1$ and we need a precision $\delta < 0.1$ for the 95% CI, the sample size required is $n > 1.96^2 \times .1 \times .9/0.1^2 = 35$. On the other hand, if we set $\alpha = 0.1$, the sample size required is $n \geq 1.65^2 \times .1 \times .9/0.1^2 = 25$. If $p^* = 0.85$, the required sample size for the 95% CI to be within $p \pm \delta$ is given by $n \geq 1.96^2 \times .85 \times .15/0.1^2 = 49$, while if we set $\alpha = 0.1$, the sample size required is: $n \geq 1.65^2 \times .85 \times .15/0.1^2 = 35$.

Since the uncertainty in (2) scales as $1/\sqrt{n}$, in the experimental design stage it may be proposed, if budget allows, to increase the sample size in order to achieve a desired precision. Based on (4), to achieve a certain accuracy in the interval estimates, say we want our $(100(1 - \alpha)\%$ confidence interval estimates of p to be within a limit δ , we just need to require that

$$z_{1-\alpha/2} \sqrt{p^*(1 - p^*)/n} \leq \delta,$$

and this results in the formula for sample size

$$n \geq z_{1-\alpha/2}^2 p^*(1-p^*)/\delta^2, \tag{6}$$

where p^* is an estimate or best guess of p .

2.3 The Wilson Interval and Bounds

Wilson interval. The *Wilson interval* (first discussed by Wilson 1927 [16]), has the form

$$\tilde{p} \pm \frac{z_{1-\alpha/2}}{\sqrt{n}} \frac{\sqrt{\hat{p}(1-\hat{p}) + t^2/4}}{1+t}, \tag{7}$$

where $\tilde{p} = (\hat{p} + t/2)/(1+t)$, and $t = z_{1-\alpha/2}^2/n$.

In (7), \tilde{p} can be considered a "shrinkage" form of \hat{p} , i.e. a weighted average between \hat{p} and $1/2$, and its standard deviation $\sqrt{\hat{p}(1-\hat{p}) + t^2/4}/(1+t)$ is a weighted and "stabilized" average form between $\hat{p}(1-\hat{p})$, and $(1/2) * (1/2)$ for $Bin(1, 1/2)$. where $Bin(n, 0.5)$ denotes the Binomial model to be defined in Section 3.1. Instead of approximating (2) by substituting the p with \hat{p} as in Wald intervals, the Wilson interval uses the theoretical standard deviation (2) and solves the endpoints: p_0 as solutions to the quadratic equation

$$(\hat{p} - p_0)^2 = p_0(1 - p_0)z_{1-\alpha/2}^2/n,$$

so Wilson intervals may improve Wald intervals in some situations and are sometimes preferred.

Example 4 Wilson Intervals and Bounds.

Suppose in two experiments when have data $n = 20, s = 2$ and $n = 20, s = 17$, and $\hat{p} = 2/20 = 0.1$ and $\hat{p} = 17/20 = .85$. Note that $z_{1-0.05/2} \approx 1.96$, and $t = 1.96^2/20 \approx 0.19$, so the Wilson point estimate $\tilde{p} = (0.1 + .19/2)/(1 + .19) = 0.16$ and $\tilde{p} = (0.85 + .19/2)/(1 + .19) = 0.79$. The 95% Wilson confidence intervals are: $0.16 \pm \frac{1.96}{\sqrt{20}} \frac{\sqrt{0.1 \times (1-0.1) + .19^2/4}}{1+0.19} \approx$

$$[0.04, 0.28] \text{ and } 0.79 \pm \frac{1.96}{\sqrt{20}} \frac{\sqrt{0.85 \times (1-0.85) + .19^2/4}}{1+0.19} \approx [0.66, 0.93].$$

To obtain the 95% confidence lower bound and upper bound for p after observing $s = 17$, we have $t = 1.645^2/20 = 0.1353$, so $\tilde{p} = (0.85 + .135/2)/(1 + .135) = 0.808$, the 90% confidence interval:

$$0.808 \pm \frac{1.645}{\sqrt{20}} \frac{\sqrt{0.85 \times (1 - 0.85) + .135^2/4}}{1 + 0.135} \approx [0.67, 0.91],$$

which gives the lower bound 0.67 and upper bound 0.91.

Wilson Upper and Lower Bounds. The $100(1 - \alpha)\%$ Wilson upper bound is:

$$p \leq \tilde{p} + \frac{z_{1-\alpha}}{\sqrt{n}} \frac{\sqrt{\hat{p}(1-\hat{p}) + t^2/4}}{1+t}$$

and the $100(1 - \alpha)\%$ Wilson lower bound is:

$$p \geq \tilde{p} - \frac{z_{1-\alpha} \sqrt{\hat{p}(1-\hat{p}) + t^2/4}}{\sqrt{n} (1+t)}.$$

Again, note that the $100(1 - \alpha)\%$ lower bound and upper bound correspond to the limits of the $100(1 - 2\alpha)\%$ confidence interval.

Note that, one should be careful in calculating the Wilson interval or bound, as every time α is changed, $t = z_{1-\alpha}^2/n$ should be updated as well as \tilde{p} which depends on t .

2.4 Agresti-Coull Interval and Bounds

A simple alternative to the Wilson interval which is as easy to calculate as the Wald interval is the Agresti-Coull interval: We simply replace in Wald interval (4) \hat{p} by $\tilde{P} = (s + z_{1-\alpha/2}^2/2)/(n + z_{1-\alpha/2}^2)$, and since $z_{.975} = 1.96 \approx 2$, and this is defined as $\tilde{p} = (s + 2)/(n + 4)$. The recommended interval

$$\tilde{p} \pm z_{1-\alpha/2} \sqrt{\tilde{p}(1-\tilde{p})/\tilde{n}}, \tag{8}$$

where $\tilde{n} = n + z_{\alpha}^2$ or $n + 4$, is called *Agresti and Coull interval* ([1], [2]).

The $100(1 - \alpha)\%$ upper bound is:

$$p \leq \tilde{p} + z_{1-\alpha} \sqrt{\tilde{p}(1-\tilde{p})/\tilde{n}},$$

and the $100(1 - \alpha)\%$ lower bound is:

$$p \leq \tilde{p} - z_{1-\alpha} \sqrt{\tilde{p}(1-\tilde{p})/\tilde{n}}.$$

Example 5 Agresti-Coull intervals.

Continue the two experiments when have data $n = 20, s = 2$ and $n = 20, s = 17$. We have, at $\alpha = 0.05, z_{\alpha/2} = 1.96 \approx 2$, point estimate $\tilde{p} = (2 + 2)/(20 + 4) = 0.17$ and $\tilde{p} = (17 + 2)/(20 + 4) = 0.79$. The 95% Agresti-Coull confidence intervals for the two examples are calculated as $0.17 \pm \sqrt{.17 \times .83/(20 + 4)} \approx [0.02, 0.32]$ and $0.79 \pm \sqrt{.79 \times .21/(20 + 4)} \approx [0.63, 0.95]$.

2.5 Poisson Adjustment For Rare Events

The Poisson model $Pois(np)$ is a good approximation when the observed number of successes $S = k$ is very small relative to sample size n , see Section 3.3. When the observed value $S = k$ is very small, the Poisson adjustment procedure for $100(1 - \alpha)\%$ confidence interval for p is given by

$$\frac{1}{2n} \chi^2(\alpha; 2k) \leq p \leq \frac{1}{2n} \chi^2(1 - \alpha/2; 2k + 2) \tag{9}$$

where $\chi^2(\alpha; m)$ is the upper quantile function of the chi-squared distribution with m degree of freedom. Note that $\chi^2(\alpha; m)$ can easily be found through standard software packages or existing statistics tables, such as *qchisq* function in R [14]. Table 1 lists select values for the 5% and 95% quantiles for $k = 1, 2, 3, 4, 5$. The use of χ^2 quantiles in (9) for Poisson parameter is purely for computational convenience and is based on a relation (16).

k	$\chi^2(\alpha/2; 2k)$	$\chi^2(1 - \alpha/2; 2k + 2)$
1	0.1025866	9.487729
2	0.7107230	12.591587
3	1.6353829	15.507313
4	2.7326368	18.307038
5	3.9402991	21.026070

Table 1: Tables of Select Quantiles of χ^2 Distribution For $\alpha = 0.1$

Furthermore, when observed number of successes s is close to n , similar Poisson adjustment is recommended. One can apply the *same adjustment* by using $n - s$ being small relative to n and constructing an interval or bounds for $1 - p$ and then reverting to get an interval or bounds for p , as shown in the following example.

Example 6 Poisson Adjustment For Rare Events.

When we have data $n = 20$ and $s = 2$, the 90% confidence interval for p based on (9) and Table 1 is given by

$$(0.7107230, 12.591587)/(2 \times 20) \approx (0.02, 0.31).$$

For $n = 20, s = 17$, one can apply similar adjustment by working with $n - s = 20 - 17 = 3$, which is our k value, and the confidence interval for $(1 - p)$ is given by

$$(1.6353829, 15.507313)/40 \approx (0.04, 0.39)$$

and so the 90% confidence interval for p is given by $(1 - .3876, 1 - 0.0409) = (0.61, 0.96)$. Since the 90% confidence lower and upper limits correspond to the 95% confidence lower bounds and upper bounds, the latter are 0.61, 0.96 respectively.

3. Probabilistic Bases and Interpretations of Interval Estimation

In this section, we describe the probabilistic mechanisms that underpin the probabilistic statements for confidence intervals and confidence bounds that are covered in Section 2. The underlying data models such as binomial model and its normal or Poisson approximations when the sample size is large are discussed. The coverage probability of interval or bound estimates should be evaluated together with the need for precision and accuracy in order to provide a good basis for testing and decision making. In practice, considerations of specific situations and some prior knowledge of the testing problem other than interval estimation from testing data alone may also be important.

3.1 Binomial Model

Under most experimental testing conditions, the correct number of detection or identification S_n out of n tests can be considered to follow the binomial model, $Bin(n, p)$, with probability given by

$$P(S_n = k|p) = \binom{n}{k} p^k (1-p)^{n-k}, \quad (10)$$

where k takes any of the possible values from $\{0, 1, 2, \dots, n\}$ and p is the probability of detection or identification (for use of binomial model in non repeated trials, see [9]). Based on (10), the probability of S_n in an interval $[a, b]$ where a, b are integers $0 \leq a \leq b \leq n$ is given by the sum

$$P(a \leq S_n \leq b) = \sum_{k=a}^b \binom{n}{k} p^k (1-p)^{n-k}. \quad (11)$$

A special case of the interval probability (11) is the cumulative distribution function (CDF) $F(L)$ defined by the probabilities with $a = 0$ and $b = s$ varies from 0, 1, 2 through n , as in

$$F(s) = P(S_n \leq s) = \sum_{k=0}^s \binom{n}{k} p^k (1-p)^{n-k}. \quad (12)$$

The binomial model has mean $\mu = np$ and standard deviation (std) $\sigma = \sqrt{np(1-p)}$.

s	14	15	16	17	18	19	20
$P(S_n = s)$	0.000	0.002	0.013	0.060	0.189	0.377	0.358
$F(s) = P(S_n \leq s)$	0.000	0.003	0.016	0.075	0.264	0.642	1.000

Table 2: Probability table of CDF $Bin(20, 0.95)$

As an example, Table 2 shows probabilities of $Bin(20, 0.95)$ for select values $L = 14$ through 20. One can see the probability of this binomial distribution is concentrated around $20 \times 0.95 = 19$ and the spread is $\sigma = \sqrt{20 \times .95 \times (1 - .95)} = 0.9747 \approx 1$.

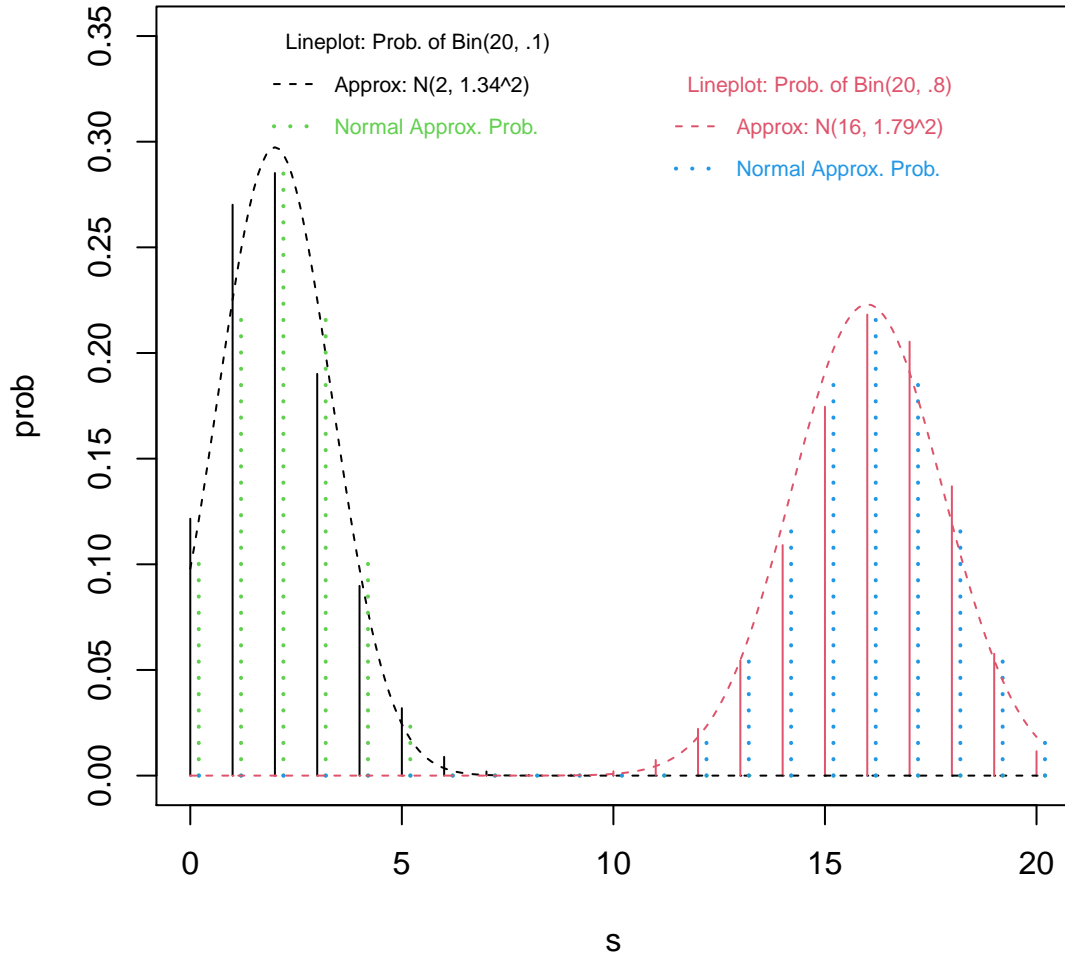


Figure 1: Display of Binomial Probabilities and Normal Approximations.

For $Bin(n, p)$ when $n = 20$ and two scenarios $p = 0.1$ (black line plot) and $p = 0.8$ (red line plot). The dashed curves are approximating normal density curves and the dotted vertical lines represent the normal probability approximations to the binomial probabilities.

3.2 Normal Approximation

The binomial distribution of S is approximated by the normal distribution $N(\mu, \sigma^2)$ with mean $\mu = np$ and standard deviation $\sigma = \sqrt{np(1-p)}$, when n is large, and p is not too close to 0 or 1, based on the Central Limit Theorem (Appendix A). That is, the interval

probability (11) from the binomial distribution is approximated by:

$$P(a \leq S_n \leq b) \approx \Phi\left(\frac{b+0.5-\mu}{\sigma}\right) - \Phi\left(\frac{a-0.5-\mu}{\sigma}\right), \quad (13)$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal $N(0, 1)$, and

$$\Phi(y) = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx. \quad (14)$$

Here 0.5 in (13) is introduced to correct for continuity as both a, b are integers and the left side of (13) is defined as in (11) while the right hand side is defined in integral form as in (14).

The normal approximation (13) says that as n is large, we can approximate a sum from a discrete probability distribution (11) by the familiar normal distribution, which is much simpler to handle.

Figure 1 illustrates the normal approximation and continuity correction for two models $Bin(20, 0.1)$ and $Bin(20, 0.8)$ where the line plots (vertical solid lines) denote the discrete binomial probabilities at given $s = 0, 2, \dots, 20$ and the dashed lines denote the approximating density curves of $N(2, 1.34^2), N(16, 1.79^2)$ (where $\mu_1 = 20 \times 0.1 = 2, \sigma_1 = \sqrt{20 \times 0.1 \times (1 - 0.1)} \approx 1.34$ and $\mu_2 = 20 \times 0.8 = 16, \sigma_2 = \sqrt{20 \times 0.8 \times (1 - 0.8)} = 1.79$). The normal approximate probabilities at each point s using (13) are shown for both cases (they are denoted by the dotted vertical lines, shifted slightly to right in order to highlight the difference). It is seen that the accuracy of normal approximation is not as good when p is close to the edge (0 or 1) as the binomial distribution is more skewed while when the binomial distribution is more symmetric as p is away from 0 or 1 and the normal approximation improves.

3.3 Poisson Approximation

When p is close to 0 or 1, the normal approximation to the Binomial distribution is not very accurate. Instead, the Poisson distribution $Pois(\mu)$ is a more accurate model where $\mu = np$, and n is the sample size. That is, when S_n follows $Bin(n, p)$, the following approximation holds

$$P(S_n = k) \approx \frac{\mu^k}{k!} e^{-\mu}, \text{ where } \mu = np, \text{ for } k = 0, 1, \dots, n, \quad (15)$$

where n is large and $\mu = np$ is assumed to be some fixed number.

The Poisson approximation $Pois(\mu)$ has mean expectation $\mu = np$ and variance $\mu = np$. So it is only close to the $Bin(n, p)$ in its first two moments (which has mean $\mu = np$ and variance $np(1 - p)$) when p is close to 0. When p is very close to 1, one can consider Poisson approximation to $n - S_n$, the total number of failures, in order to have a more accurate model for S_n .

The CDF of $Pois(\mu)$ has the curious but convenient relation

$$P(Y \leq k) = P(\chi_{2k+2}^2 \geq 2\mu), \quad (16)$$

s	Binom	Normal	Poisson
0	0.12	0.13	0.14
1	0.39	0.35	0.41
2	0.68	0.64	0.68
3	0.87	0.87	0.86
4	0.96	0.97	0.95
5	0.99	0.99	0.98
6	1.00	1.00	0.99

Table 3: Approximation of CDF $Bin(20, 0.1)$ by $N(\mu - 0.5, \sigma^2)$ and $Pois(\lambda)$, Where $\mu = \lambda = 20 \times 0.1 = 2$, $\sigma^2 = 20 \times .1 \times .9 = 1.8$

where Y follows $Pois(\mu)$ and χ_m^2 is the standard χ_m^2 distribution with m degree of freedom. (16) is used in constructing a confidence interval in Subsection 2.5.

As an example, Table 3 shows the comparison in CDFs of $Bin(n, p)$ by normal approximation $N(\mu, \sigma^2)$ and the Poisson distribution $Pois(\mu)$ where $\mu = np$, $\sigma^2 = np(1 - p)$ and $n = 20$, $p = 0.1$.

3.4 Asymptotic Coverage and Relation to Testing

Interval construction for p may be defined based on either using the exact probability (11), see Section 4, or large sample approximation such as (13). The normality approximation gives rise to a simple and easy to compute approximate procedure for confidence intervals.

If we choose a, b so that the right side of (13) is of probability $(1 - \alpha)$ for some $0 < \alpha < 1$, for example,

$$(b + 0.5 - \mu)/\sigma = z_{1-\alpha/2}, (a - 0.5 - \mu)/\sigma = z_{\alpha/2},$$

where z_c denotes the quantile function of standard normal such that

$$\int_{-\infty}^{z_c} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = c, \text{ for } 0 < c < 1,$$

and by symmetry of normal distribution, $z_c = -z_{1-c}$. Then, we can say that

$$P(np - 0.5 - z_{1-\alpha/2}\sqrt{np(1-p)} \leq S_n \leq np + 0.5 + z_{1-\alpha/2}\sqrt{np(1-p)}) \rightarrow 1 - \alpha. \quad (17)$$

Interval estimation for p can be defined based on (17), leading to Wald interval and Wilson interval, previously presented in subsection 2.1 and subsection 2.3.

From Equation (17), a procedure for testing hypothesis $H_0 : p = p_0$ against the alternative $H_a : p \neq p_0$ when the testing procedure based on S is: reject H_0 when p_0 does not belong to the confidence intervals based on S . The test has significance level α , when the $(1 - \alpha)$ confidence interval is used as a testing procedure.

A testing procedure for one-sided test of $H_0 : p \leq p_0$ vs $H_a : p > p_0$ based on observed data S is given by: reject H_0 if p_0 is less than the confidence lower bound of p . For testing $H_0 : p \geq p_0$ vs $H_a : p < p_0$, reject H_0 if p_0 is bigger than the confidence upper bound.

As an example, suppose we want to evaluate whether a testing system has 95% or above probability of detection rate. This can be formulated as a hypothesis testing problem $H_0 : p \geq 0.95$ vs $H_a : p < 0.95$. That is, our goal is to establish and reject an inferior product (to accept $H_a : p < 0.95$) if there is sufficient evidence to warrant it. We should expect to observe an S which is closer to n and so in order to apply the Poisson adjustment procedure (9), we consider data $k = n - s$ which should follow $Bin(n, 1 - p)$. the Poisson adjusted interval for $1 - p$ is given by

$$\frac{1}{2n} \chi^2(\alpha; 2k) \leq 1 - p \leq \frac{1}{2n} \chi^2(1 - \alpha/2; 2k + 2). \quad (18)$$

Set $\alpha = 0.1$, a test of $H_0 : p \geq 0.95$ (or $1 - p \leq 0.05$) is given by: rejecting H_0 if $(1 - p_0)$ if outside of the bound provided by (18), that is, if $1 - p_0$ is less than the lower limit, or

$$\chi^2(\alpha; 2k) > 2n(1 - p_0). \quad (19)$$

Example 7 Relating Confidence Bound to Testing Decision.

For a testing experiment with $n = 20, p_0 = 0.95$, we need $\chi^2(\alpha; 2k) > 2 \times n \times (1 - 0.95) = 2$ from (19). From Table 1, k is required to be bigger or equal to 4 $k = n - s \geq 4$, and so $s \leq 20 - 4 = 16$ is the criterion for this experiment to reject $H_0 : p \geq 0.95$ at the significance level $\alpha = 0.05$. That is, in other words, if we obtain $s \geq 17$ out of $n = 20$ tests, we cannot reject $p \geq (p_0 = 0.95)$ while if $s \leq 16$, we conclude that $p < p_0 (= 0.95)$.

From Example 7, the decision point of $s = 17$ out of $n = 20$ appears to be an interesting decision point, as from Table 2, if $p = 0.95$ is true, there is only 0.016 chance that $S \leq 16$ is observed! On the other hand, if one adopts this decision criterion, one also lives with the consequence that whether we accept too many instruments which may not have met the performance target. Indeed. if we calculate $F(16)$ of $Bin(20, p)$ as in Table 2 for other p values, say $p = 0.7, 0.8, .9$, we obtain $F(16) \approx 0.89, 0.59, 0.13$ so that we have probability $1 - F(16)$: 0.11, 0.41, 0.87 of accepting $p = 0.7, 0.8, 0.9$ of meeting the performance target of $p \geq 0.95$.

4. Exact Interval and Bayesian Intervals

Confidence intervals and bounds as discussed in Section 2 are widely used and since they are so easy to calculate and use, we recommend their use, except when the sample size n is small. When the **sample size n is small**, these methods tend to give unreasonable results. A wide class of methods based on the exact Binomial model of Subsection 3.1, which can be put in the general Bayesian framework will be discussed in this section, give more natural results and appealing alternatives. First is the classic Clopper-Pearson interval or exact methods, which is discussed in Section 4.1. The rest of this Section deal with the general Bayesian formulation which represents another class of inference tools which, not only allow us to address estimation or inference of probability P_d or P_{ID} directly (instead of being based on sampling theory of estimators), but also allow incorporation of prior information and different sources of information on the parameter of interest into the current testing problem.

4.1 Clopper-Pearson Interval

When the sample size n is small to moderate, the Binomial model $Bin(n, p)$ plays a more prominent role in relating data to unknown parameter, both as a forward model for data S and as a basis for inference on the parameter p .

Clopper-Pearson "Exact" Interval. The Clopper-Pearson [7] "exact" confidence interval for p , based on inverting equal-tailed binomial test of $H_0 : p = p_0$ based on the Binomial model $Bin(n, p)$ for data S .

The exact interval $[p_1(S), p_2(S)]$ have endpoints that are the solutions in p_0 to the equations

$$\sum_{k=s}^n \binom{n}{k} p_0^k (1-p_0)^{n-k} = \alpha/2$$

and

$$\sum_{k=0}^s \binom{n}{k} p_0^k (1-p_0)^{n-k} = \alpha/2$$

except that the lower bound is 0 when $S = 0$ and the upper bound is 1 when $S = n$, where s is the observed value of S . Note that the interval so constructed satisfies

$$P(p_1(S) \leq p \leq p_2(S)) = 1 - \alpha \tag{20}$$

for every $p : 0 < p < 1$, where the above coverage probability is evaluated under the model $S \sim Bin(n, p)$.

From Subsection 5.3, the lower endpoint $p_1(s)$ is the $\alpha/2$ quantile of a Beta distribution with parameters s and $n - s + 1$, or $g_{\alpha/2, s, n-s+1}$ in notation defined in (32), and the upper endpoint $p_2(s)$ is the $(1 - \alpha/2)$ quantile of a Beta distribution with parameters $s + 1$ and $n - s$, or $g_{1-\alpha/2, s, n-s+1}$.

In words, the confidence statement of (20) says that if the same experiments (with same sample size n) were repeated many times, resulting in different s 's, from which the

confidence limits are computed each time, the fraction of times that the confidence interval contains the true p is exactly $(1 - \alpha)$, or bigger.

Example 8 Clopper-Pearson Exact Interval.

Suppose we have in two instances of observed data: $n = 20, s = 2$ and $n = 20, s = 17$. The 95% "exact" Clopper-Pearson interval in the first case, is given by $q_{0.025, 2, 20-2+1} \approx 0.012$ and $q_{0.975, 2+1, 20-2} \approx 0.317$ and so is $[0.012, 0.317]$. And in the second case, endpoints are given by $q_{0.025, 17, 20-17+1} \approx 0.621$ and $q_{0.975, 17+1, 20-17} \approx 0.968$, and so is $[0.621, 0.968]$. Similarly, the 90% confidence interval is given by $[q_{0.05, 17+1, 20-17}, q_{0.95, 17+1, 20-17}] \approx [0.656, 0.929]$, and the 80% confidence interval is given by $[q_{0.10, 17+1, 20-17}, q_{0.90, 17+1, 20-17}] \approx [0.696, 0.910]$. Note that the 95% lower and upper bounds correspond to the limits of the 90% confidence interval, that is, 0.656, 0.929. And the 90% lower and upper bounds correspond to the limits of the 80% confidence interval, that is, 0.696, 0.910.

4.2 Bayesian Inference

Bayesian inference provides a direct approach to estimation and inference of probability of detection, which embodies uncertainty as well as prior information in the form of posterior distribution, on which testing and decision based on heterogeneous sources of information can be made in a straightforward way. The only obstacle, in addition to some calculations involved, is the specification of prior information on the probability of interest. The prior information may be based on historical and earlier experiences, subjective opinion, or some "uninformative" priors as the last resort. A simple recipe for implementing these procedures is provided through the Beta prior family and calculations as well as comparison to earlier approaches are given in this section.

Data model: With limited resources, the n test should be distributed over many possible scenarios and so the test results X_1, \dots, X_n can have different probabilities of detection, that is, $P(X_i = 1) = p_i, P(X_i = 0) = 1 - p_i$, and if one can assume that $(X_1, p_1), \dots, (X_n, p_n)$ are independent and *exchangeable*, and p_i 's are from some distribution $\pi(p), 0 < p < 1$, then the sum $S = \sum_{i=1}^n X_i$, the total number of successes in n tests, obeys the model:

$$P(S = s) = \binom{n}{s} \int_0^1 p^s (1-p)^{n-s} \pi(p) dp, s = 0, 1, \dots, n \quad (21)$$

where $\pi(\cdot)$ is some probability density function on the interval $[0, 1]$, and is called the *prior distribution*.

Bayesian inference: For now, what can be said about p based on observed data $S = s$ which comes from the model (21). It turns out that by a simple application of the Bayes Theorem in probability theory [4], the distribution of p after observing data $S = s$ can be given explicitly:

$$P(p|s) = \frac{p^s (1-p)^{n-s} \pi(p)}{\int_0^1 p^s (1-p)^{n-s} \pi(p) dp}, 0 < p < 1 \quad (22)$$

which is called the posterior distribution, containing all inferential information on p .

Likelihood function: After observing data $S = s$, all *information about the data* (n, s) is contained in the *likelihood function*

$$L(p|s) = \binom{n}{s} p^s (1-p)^{n-s} \propto p^s (1-p)^{n-s}, 0 < p < 1, \quad (23)$$

that is, the same binomial probability is treated as a function of p as the probability p varies from 0 to 1 continuously.

So the posterior distribution (22) has the verbal form:

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{P(\text{data})}$$

where $P(\text{data})$ is integration or sum of $\int_0^1 \text{Likelihood} \times \text{Prior} dp$ as it is simply the normalizing constant that makes the numerator integrate to 1, so any constant in front of the likelihood or the prior will be canceled out anyway and does not need to be specified exactly.

Beta prior family: A particular class of model for $\pi(p)$ is *the conjugate prior*: the Beta distribution $Be(\gamma, \beta)$ with density

$$b(p; \gamma, \beta) \propto p^{\gamma-1} (1-p)^{\beta-1}, 0 < p < 1, \gamma > 0, \beta > 0.$$

$$b(p; \gamma, \beta) = \frac{p^{\gamma-1} (1-p)^{\beta-1}}{B(\gamma, \beta)}, 0 < p < 1,$$

where $B(\gamma, \beta) = \int_0^1 p^{\gamma-1} (1-p)^{\beta-1} dp = \Gamma(\gamma)\Gamma(\beta)/\Gamma(\gamma+\beta)$. Beta prior models have a rich shape and can be used to model most prior information and subjective opinion, see [4] for a very clear and intuitive discussion.

It is both important and convenient because it leads to a very convenient posterior distribution for p , given by $Be(s+\gamma, n-s+\beta)$, which has density:

$$b(p; \gamma+s, n-s+\beta) = \frac{p^{\gamma+s-1} (1-p)^{n-s+\beta-1}}{B(\gamma+s, n-s+\beta)}, 0 < p < 1 \quad (24)$$

where

$$B(\gamma+s, n-s+\beta) = \Gamma(\gamma+s)\Gamma(n-s+\beta)/\Gamma(n+\gamma+\beta).$$

4.3 Bayesian Posterior Summaries

Bayesian inference says that everything we need to know about p is contained in the posterior distribution (22). The Bayesian principle is very similar to the likelihood principle, especially when the prior distribution is very vague or non-informative. In practice, we can in principle reporting the whole posterior distribution whenever possible; though, there may

be occasions when one may be also interested in some simple summaries of the posterior distribution, even just for comparison with the classical approaches.

The Beta posterior distribution (24) has mean

$$E[p|s] = \frac{s + \gamma}{n + \gamma + \beta} = \tilde{p}_b,$$

and variance

$$\text{var}[p|s] = \frac{(s + \gamma)(n - s + \beta)}{(n + \gamma + \beta)^2(n + \gamma + \beta + 1)} = \frac{1}{n + \gamma + \beta + 1} \tilde{p}_b(1 - \tilde{p}_b).$$

The posterior mode is given by:

$$\text{mode}[p|s] = \frac{s + \gamma - 1}{n + \gamma + \beta - 2}.$$

One can also calculate the posterior median, which is given by the 50% quantile of the Beta distribution $Be(s + \gamma, n - s + \beta)$, or $g_{0.5; s + \gamma, n - s + \beta}$ (refer to (32)). These estimators can be compared to the other estimators of p discussed in Section 2.

4.4 Prior Choice and Incorporating External Information

Specification of $\pi(p)$ is a crucial issue in Bayesian analysis. Not only one needs to specify a proper prior $\pi(p)$ so that the posterior distribution is proper, one would like to incorporate any other information into the prior information as much as possible, while also one does not want the prior information to overwhelm the test data, so we will focus on a few obvious choices which have least impact on the likelihood.

One such prior is the uniform distribution on $[0, 1]$, which corresponds to the Beta distribution with $\gamma = \beta = 1$, i.e. $Be(1, 1)$. The resulting posterior distribution of p is the Beta distribution $Be(s + 1, n - s + 1)$, which has the same form as the likelihood function,

$$P(p|s) \propto p^s(1 - p)^{n-s}$$

and so we also call the approach with uniform prior the *Bayesian likelihood method*. This will be our default method we will use in the rest of this section as it gives exactly the result based on what data alone can provide, and all other choices of prior will modify or enhance this likelihood information.

Another often recommended noninformative prior choice is the Jeffreys' prior, corresponding to $Be(\frac{1}{2}, \frac{1}{2})$, or

$$\pi(p) \propto 1/\sqrt{p(1 - p)},$$

which puts more weight toward the endpoints at $p = 0$ and $p = 1$.

To have a posterior mean that approximates the Agresti-Coull estimator $\tilde{p} = (s + z_{1-\alpha/2}^2)/(n + z_{1-\alpha/2}^2)$, which is approximately $(s + 2)/(n + 4)$ for $\alpha = 0.05$, the corresponding prior distribution is $Be(2, 2)$, that is,

$$\pi(p) \propto p(1 - p)$$

for which, for lack of name, we call it the Agresti prior. Note that the corresponding posterior distribution has mode is $(s + 1)/(n + 2) = \frac{n}{n+2} \frac{s}{n} + \frac{2}{n+2} \frac{1}{2}$, which is a shrinkage of MLE $\frac{s}{n}$ and $\frac{1}{2}$.

The three priors, $Be(1, 1)$, $Be(2, 2)$, $Be(\frac{1}{2}, \frac{1}{2})$ represent the three types of symmetric prior information on p : putting uniform weight across the whole range of $[0, 1]$, putting more weight at the center $p = 0.5$, and putting more weight equally toward the end at $p = 0$ and $p = 1$, see Figure 2. After observing binomial data n, s , the corresponding posteriors for using any of the three priors are $Be(s + 1, n - s + 1)$, $Be(s + 2, n - s + 2)$, $Be(s + 0.5, n - s + 0.5)$ respectively, based on (24). A summary of formulas for Bayesian inference under the three priors is given in Table 4.

Prior choice	Data	Likelihood	Posterior (up to a constant)
Uniform: $\pi(p) = 1$	n, s	$p^s(1 - p)^{n-s}$	$p^s(1 - p)^{n-s}$
Jeffreys: $1/\sqrt{p(1 - p)}$	n, s	$p^s(1 - p)^{n-s}$	$p^{s-0.5}(1 - p)^{n-s-0.5}$
Agresti: $p(1 - p)$	n, s	$p^s(1 - p)^{n-s}$	$p^{s+1}(1 - p)^{n-s+1}$

Table 4: Bayesian Inference of Binomial Data Under Select Prior Choices

Predictive model and model assessment: Under the uniform prior, the corresponding data model (21) is simply the uniform distribution on $\{0, 1, 2, \dots, n\}$, that is s takes on any of the $n + 1$ possible values with probability $1/(n + 1)$ (a fact apparently recognized by Thomas Bayes himself in his original 1764 paper [15]). In general, under prior $Be(\gamma, \beta)$, the data model (21) is

$$P(S = s) = \binom{n}{s} B(\gamma, \beta)^{-1} B(\gamma + s, \beta + n - s) \quad (25)$$

which is known as the *Beta-Binomial distribution*. One may use (25) to assess the choice of prior parameters γ, β for given data s . In particular, choosing the three priors implies, in addition to that the uniform prior implies equal prior chance of observing s among the possible values $\{0, 1, 2, \dots, n\}$, the Agresti prior implies that there is a higher chance of observing s in the middle, say $n/2$ or $(n + 1)/2$ than of values at the end near 0 or n , while the Jeffreys prior implies that there is a much higher prior chance of obtaining s in the edge near 0 or 1 than in the middle.

Since interest in estimation of probability of detection (or identification) is focused near the high end, between $(0.5, 1]$ while the probability of false alarms is focused in the lower end $[0, 0.3)$, we think the Jeffreys prior gives undue weight near the two edges, and so among the three priors, we favor the uniform prior first, the Agresti prior next and the Jeffreys prior the least. In the following subsections, we will focus on using the uniform prior as our default choice to illustrate the Bayesian methodology, but we still use the Jeffreys and Agresti priors for purpose of comparison and for sensitivity analysis of prior choice in Bayesian analysis.

We should point out that none of the mentioned prior choices should be the automatic choice for a given problem, except maybe the likelihood method when it results in a proper

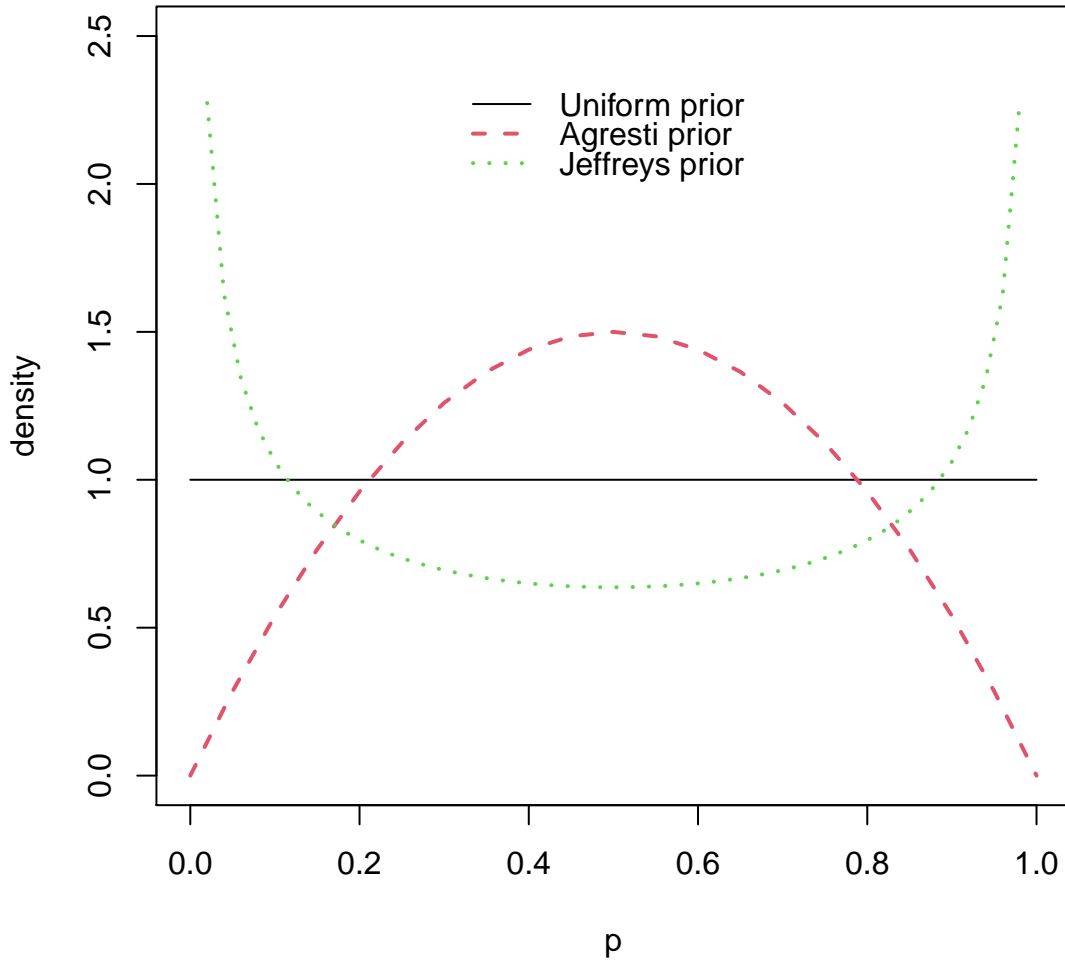


Figure 2: Display of Three Beta Distributions.

Uniform Prior ($Be(1, 1)$) (black, solid line), Agresti Prior: $Be(2, 2)$ (red, dashed line), Jeffreys Prior $Be(0.5, 0.5)$ (green, dotted line).

posterior distribution. Several broad approaches to specifying prior information are suggested by Berger ([3], Chapter 3) and we concur with his recommendation for active use of subjective specification of prior base on substantial knowledge or experience, whenever possible, and encourage use of empirical, semi-empirical (maximum entropy) methods to utilize partial information from historical or other sources of information. The use of non-informative priors should be a last resort, but choosing a particular noninformative prior

does provide a Bayesian alternative tool for providing comparison or support to interval estimation results from any of the classical approaches such as those from Section 2.

4.5 Bayesian Credible Intervals

Since the Bayesian posterior distribution contains all the information about p , there is no need for a separate theory for constructing Bayesian intervals or Bayesian testing theory. However, there are various summaries of Bayesian posterior distribution which mimic and can be compared with the classical interval procedures. One example is the equal-tailed Bayesian credible interval, which is provided by

$$[g_{\alpha/2; s+\gamma, n-s+\beta}, g_{1-\alpha/2; s+\gamma, n-s+\beta}]$$

where $g_{\alpha/2; s+\gamma, n-s+\beta}, g_{1-\alpha/2; s+\gamma, n-s+\beta}$ are the $100(\alpha/2)\%$ and $100(1 - \alpha/2)\%$ quantile of the posterior Beta distribution $Be(s + \gamma, n - s + \beta)$. This interval has the intuitively more understandable interpretation that p is contained in the given interval with probability $(1 - \alpha)$ for any given data n, s . To distinguish it from the previous classical confidence intervals, this Bayesian interval is called *credible interval*. Note that for credible interval, the probability is defined on the space of p while previously with the classical confidence intervals the (coverage) probability is on the observed data space!

In addition to the two-sided equal-tailed Bayesian credible intervals, there are, of course, other means of defining a credible interval for p that has, under the posterior distribution, probability $(1 - \alpha)$. For example, any such asymmetric tailed interval

$$[g_{\alpha/c; s+\gamma, n-s+\beta}, g_{1-\alpha*(c-1)/c; s+\gamma, n-s+\beta}]$$

for any $c > 1$ defines an $100(1 - \alpha)\%$ credible interval. One can search over different c 's to find the shortest interval, and this gives rise to an easy approximation to the highest posterior density (HPD) interval, which is, the most precise and powerful interval from Bayesian analysis.

In particular, it is easy to define and compute one-sided $100(1 - \alpha)\%$ Bayesian intervals, or **lower Bayesian Bound (LB)**, **upper Bayesian bound (UB)**. It is given by

$$LB_{\alpha}(n, s) = g_{\alpha, s+\gamma, n-s+\beta},$$

and

$$UB_{\alpha}(n, s) = g_{1-\alpha, s+\gamma, n-s+\beta}.$$

Note that [LB, UB] is the same as the two-sided equal-tailed $100(1 - 2\alpha)\%$ Bayesian interval. For example, the 95% lower and upper Bayesian bounds are given by the 90% limits of the two-sided equal-tailed Bayesian interval.

A few comments are in order: as seen from Example 9: clearly, the Jeffreys prior $Be(0.5, 0.5)$ pulls both point estimate and the interval estimate toward either of the two ends (either $p = 0$ or $p = 1$ depending on which side s is), while the Agresti prior $Be(2, 2)$ pulls estimates toward to the middle ($p = 0.5$). The uniform prior $Be(1, 1)$ gives interval

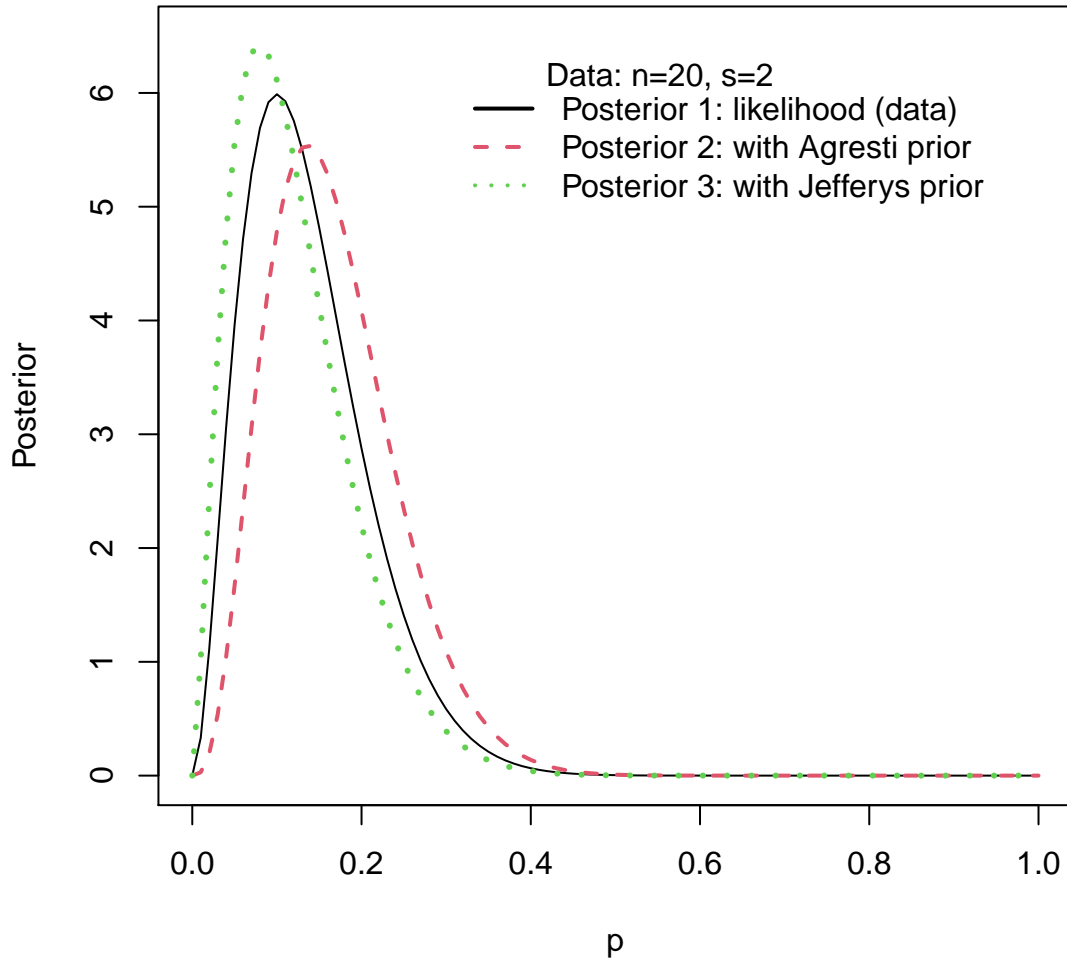


Figure 3: Illustration of Bayesian Inference: Data $n = 20, s = 2$ and Uses of Agresti and Jeffreys priors Are Compared to the Likelihood.

and point estimates between these two, and may be considered the least unbiased by use of prior information, even though the median estimates from the Jeffrey prior are closest to the classical unbiased estimates in Example 2.

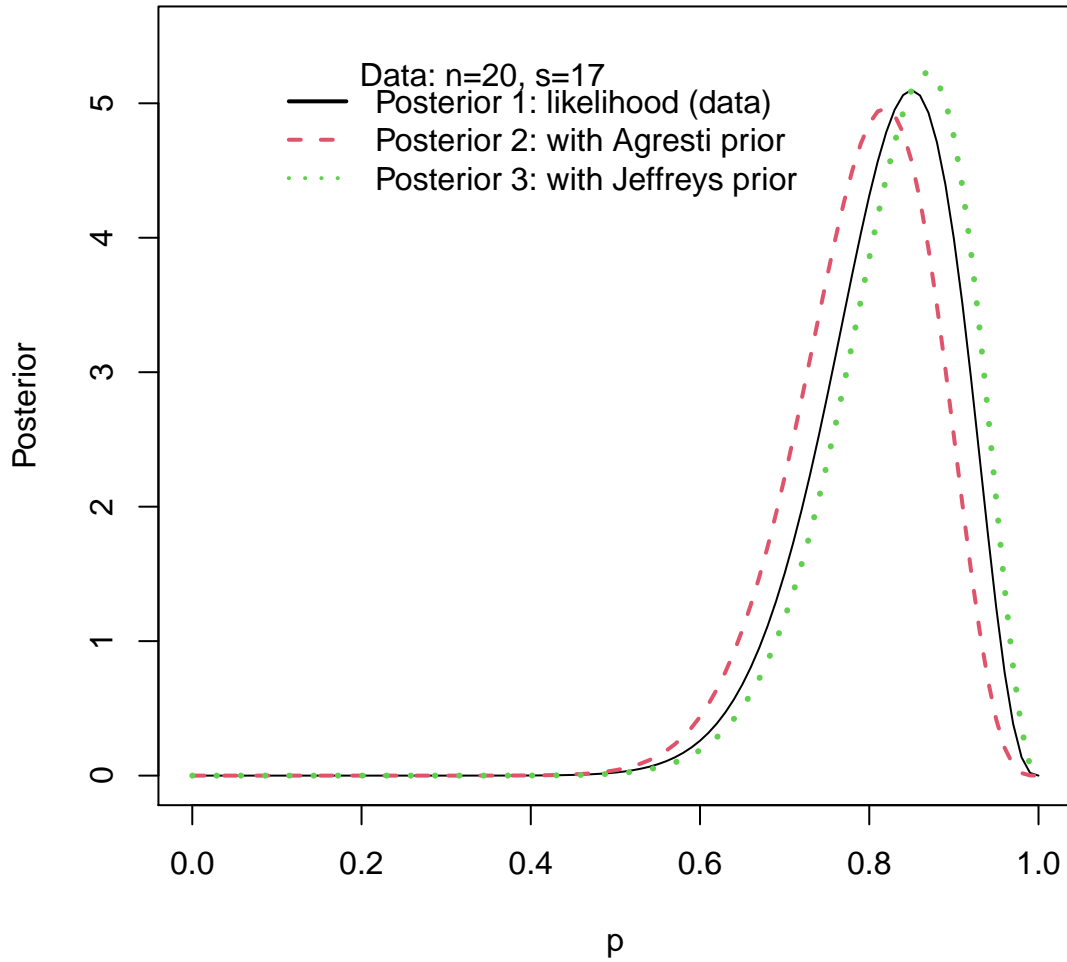


Figure 4: Illustration of Bayesian Inference: Data $n = 20, s = 17$ and Uses of Agresti and Jeffreys priors Are Compared to the Likelihood.

4.6 Comparison Among the Exact and Bayesian Intervals

The Clopper-Pearson confidence interval can be written as

$$CI_{CP} = [g_{\alpha/2;s;n-s+1}, g_{1-\alpha/2;s+1;n-s}],$$

Example 9 Bayesian and Likelihood Intervals.

Assume that we obtained binomial data in two instances, $n = 20, s = 2$ and $n = 20, s = 17$. Using each of the Beta prior distribution, $Be(1, 1), Be(2, 2), Be(0.5, 0.5)$, we can report the the corresponding posteriors $Be(2 + 1, 20 - 2 + 1), Be(2 + 2, 20 - 2 + 2), Be(2 + 0.5, 20 - 2 + 0.5)$ for the first case, shown in Figure 3 and $Be(17 + 1, 20 - 17 + 1), Be(17 + 2, 20 - 17 + 2), Be(17 + 0.5, 20 - 17 + 0.5)$ in the second case, shown in Figure 4. Evaluation of Bayesian posterior distribution under different prior choices, or *sensitivity analysis*, allows us to see the impact of prior choice on the Bayesian inference. The 95% Bayesian credible intervals (with posterior median) for both cases are given in Table 5, where the left three columns with lower limit (LL), median (Med), upper limit (UL) are for the first case $n = 20, s = 2$ and the next three columns are for $n = 20, s = 17$. The three rows correspond to the choice of three priors.

Prior	Posterior	$n = 20 \quad s=2$			$n=20 \quad s=17$			
		LL	Med	UL	LL	Med	UL	
Be(1,1)	Be(3, 19)	0.030	0.125	0.304	Be(18, 4)	0.637	0.828	0.946
Be(2,2)	Be(4, 20)	0.050	0.157	0.336	Be(19, 5)	0.612	0.800	0.925
Be(.5, .5)	Be(2.5, 18.5)	0.021	0.107	0.284	Be(17.5, 3.5)	0.651	0.844	0.956

Table 5: 95% Bayesian Credible Intervals $[LL, UL]$ (with median value Med) Under Three Prior Choices for Observed Data $n = 20, s = 2$ (columns 2,3,4) and $n = 20, s = 17$ (columns 5,6,7)

where $g_{c,\gamma,\beta}$ is the $100c\%$ quantile of the Beta distribution $Be(\gamma, \beta)$ for any $0 < c < 1$ (See (32)). This is very close to the Bayesian credible interval with the Jeffreys prior

$$CI_J = [g_{\alpha/2;s+0.5,n-s+0.5}, g_{1-\alpha/2;s+0.5,n-s+0.5}].$$

It follows that CI_{CP} always contain CI_J , and CI_{CP} is considered too conservative and CI_J corrects it to some extent, a so-called Mid-P Clopper-Pearson is also proposed to replace CI_{CP} by Brown, Cai, DasGupta ([6], p.115). Figure 5 illustrates the comparison of interval estimates based on the Clopper-Pearson, Jeffreys, and likelihood methods for all possible data $s = 0, 1, 2, \dots, 20$ when $n = 20$.

Among the Bayesian interval estimates, in addition to the Jeffreys interval, the interval based on the Agresti prior $Be(2, 2)$, given by

$$CI_A = [g_{\alpha/2;s+2,n-s+2}, g_{1-\alpha/2;s+2,n-s+2}]$$

pulls estimates toward the middle $p = 0.5$, which may be too much. The Bayesian interval based on the uniform prior $Be(1, 1)$, which we may call the Bayesian likelihood interval, as it is exactly the same as that given by the Binomial likelihood,

$$CI_A = [g_{\alpha/2;s+1,n-s+1}, g_{1-\alpha/2;s+1,n-s+1}].$$

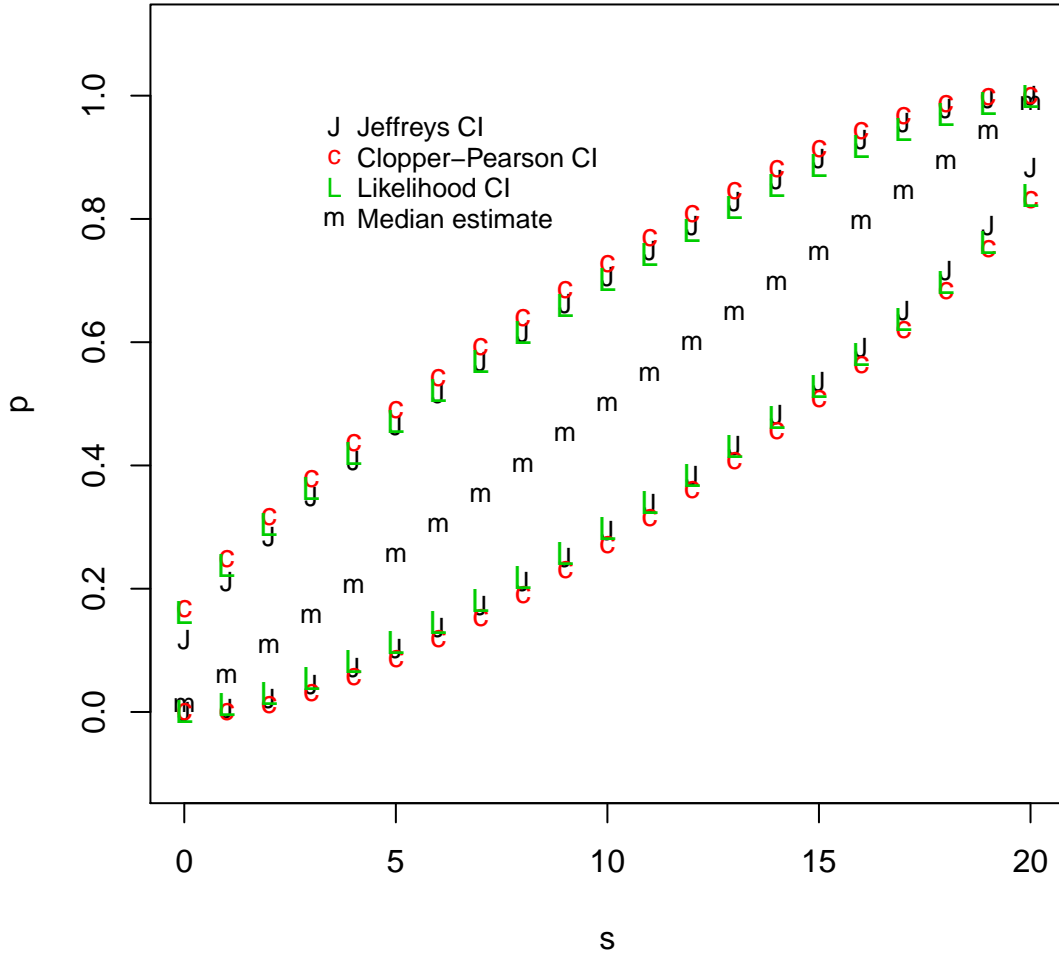


Figure 5: Comparison of Interval Estimates Using Clopper-Pearson Exact Method, Jeffreys Credible Intervals and Bayesian Likelihood Intervals.

Including median estimate from the Jeffrey prior, with $n = 20$ given at every potential data value $s = 0, 1, 2, 3, \dots, 20$ labeled in x-axis.

Figure 5 illustrates the comparison of Bayesian credible intervals for Jeffreys, uniform, and Agresti priors for all possible data $s = 0, 1, 2, 3, \dots, 40$ when $n = 40$.

It is observed that in general, when data s is in the middle range further away from 0 or n , all intervals are very close to each other. Jeffreys prior pulls interval estimates too aggressively toward the edge $p = 0$ or 1 when s is close to 0 or n , and it is recommended

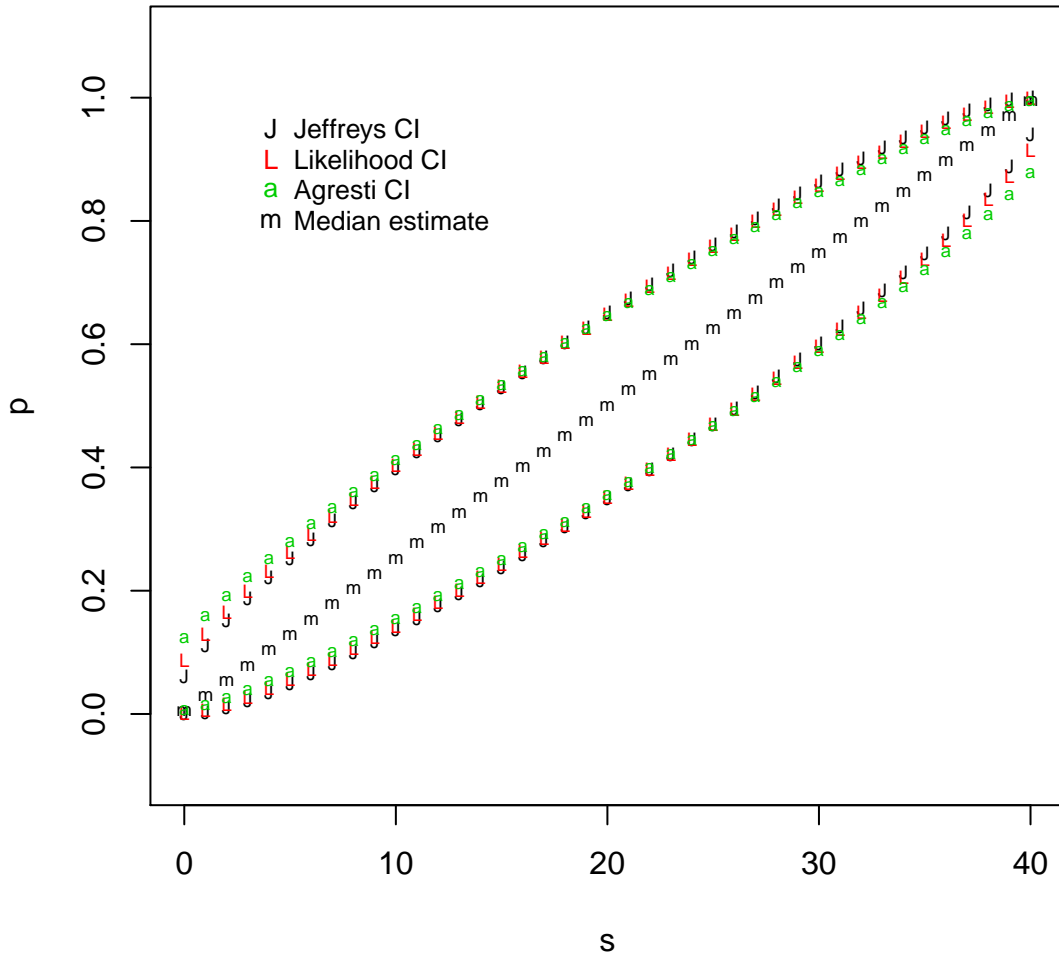


Figure 6: Comparison of Bayesian Credible Intervals: with Jeffreys, Uniform, and Agresti Priors

With $n = 40$, given at every potential data $s = 0, 1, 2, \dots, 40$ labeled in x-axis. The median estimate is computed based on the posterior using the Jeffreys prior.

that some adjustment on Jeffreys interval needed when p is very close to 0 and 1 (Section 4.1 of Brown, Cai, DasGupta[6]). On the other hand, the Agresti prior pulls estimates more to $p = 0.5$ when s is close to either end of 0 or n and so it is too biased toward $p = 0.5$. Thus, we recommend the use of the Bayesian credible interval with uniform prior, or what we call the *Bayesian likelihood interval* as our first choice, as it gives the most unbiased

results in the sense of least influence from the use of a prior distribution.

Based on the calculation method from Subsection 5.3, Table 6, Table 7, Table 8 list the 95% limits of the Bayesian likelihood intervals for select sample size $n = 7$ up to 30. We note that Brown, Cai and DasGupta [6] already provided two useful tables for 95% limits of the Jeffreys intervals for sample size n from 7 through 30.

5. Statistical Analysis of Quantitative Data

When collection of actual performance data is feasible and is planned ahead in a testing experiments, more statistical tools become available for estimation of probability of detection or identification, in addition to other useful purposes. Both parametric model and nonparametric approaches will be discussed in this section. For the purpose of probability estimation, the nonparametric approach is comparable to the binary data analysis discussed in previous sections, in terms of sample size requirements and accuracy of confidence intervals, the parametric approach is new and is based on the assumption that a parametric model can be found and be fitted to the quantitative performance data. The parametric probability estimation approach is known to be more accurate and requires less data provided that the underlying parametric model is adequate.

5.1 Confidence Intervals for Parameters

Confidence Intervals For the Mean.

Given test data x_1, x_2, \dots, x_n as we discussed in Subsection 1.3, the sample mean \bar{x} provides a reasonable estimate for the true mean because of the "central limit theorem" (Subsection 5.3) when the sample size n is large enough. Thus, a $100(1 - \alpha)\%$ confidence interval for μ is given by

$$\mu \in [\bar{x} \pm z_{1-\alpha/2} \frac{D}{\sqrt{n}}]. \quad (26)$$

This interval is valid in the sense of having the right coverage probability of $1 - \alpha$ only for large n . The 95% confidence lower bound (LB) and upper bound (UB) are given by

$$\mu \geq \bar{x} - z_{1-\alpha} \frac{D}{\sqrt{n}},$$

and

$$\mu \leq \bar{x} + z_{1-\alpha} \frac{D}{\sqrt{n}}.$$

Note the difference in the critical factor $z_{1-\alpha/2}$ from the two-sided confidence interval to $z_{1-\alpha}$ for the one-sided bounds. So the lower bound and upper bound $[LB, UB]$ are the same as the $100(1 - 2\alpha)\%$ limits of the two-sided confidence interval.

Confidence Interval for the Mean: Unknown Variance.

Since D is only an estimate of the true σ , the standard deviation of the population, the interval estimation (26) can be improved to take into account this additional uncertainty. When the normal distribution can be assumed, that is, the underlying data x_1, x_2, \dots, x_n can be considered as coming from a normal distribution $N(\mu, \sigma^2)$, an exact two-sided $100(1 - \alpha)\%$ confidence interval for the mean μ can be defined, it is given by:

$$\mu \in [\bar{x} \pm t_{1-\alpha/2, n-1} \frac{D}{\sqrt{n}}]. \quad (27)$$

where $t_{1-\alpha/2,n}$ denotes the upper $(1 - \alpha/2)$ th quantile of the Student t -distribution with $n - 1$ degree of freedom. The $100(1 - \alpha)\%$ one-sided lower confidence bound (LB) and upper confidence bound (UB) for μ are given by:

$$\mu \geq \bar{x} - t_{1-\alpha,n-1} \frac{D}{\sqrt{n}},$$

and

$$\mu \leq \bar{x} + t_{1-\alpha,n-1} \frac{D}{\sqrt{n}}.$$

Confidence Interval for Standard Deviation.

The sample standard deviation D has a simple distribution if the underlying data come from the normal population $N(\mu, \sigma^2)$. One can show that $(n - 1)D^2$ follows $\sigma^2 \chi_{n-1}^2$, where χ_{n-1}^2 is the Chi-squared distribution with $(n - 1)$ degree of freedom. So the $100(1 - \alpha)\%$ confidence interval for σ^2 is given by

$$\frac{(n - 1)D^2}{\chi^2(1 - \alpha/2, n - 1)} \leq \sigma^2 \leq \frac{(n - 1)D^2}{\chi^2(\alpha/2, n - 1)},$$

and the $100(1 - \alpha)\%$ confidence interval for σ is given by

$$\sqrt{\frac{(n - 1)D^2}{\chi^2(1 - \alpha/2, n - 1)}} \leq \sigma \leq \sqrt{\frac{(n - 1)D^2}{\chi^2(\alpha/2, n - 1)}},$$

where $\chi^2(c; n - 1)$ is the c upper quantile function of the chi-squared distribution with $n - 1$ degree of freedom.

Example 10 Confidence Intervals For Parameters.

Suppose we have an experiment with test results of 15 measurements of the time to detection which was measured in seconds: 91, 95, 107, 105, 102, 85, 88, 92, 101, 99, 102, 85, 114, 91, 95. This gives $n = 15, \bar{x} = 96.8(s), D = 8.44$. With $t_{.95,14} = 1.761$, the 90% confidence interval for the mean based on (27) is given by

$$96.8 \pm 1.761 \times \frac{8.44}{\sqrt{15}} = [92.96, 100.64].$$

Note that $t_{.95,14}$ is only slightly larger than $z_{0.95} = 1.645$ so the t -based interval is slightly wider than the normal based interval (26). Find $\chi^2(.05, 14) = 6.57, \chi^2(.95, 14) = 23.65$, the 90% confidence interval for σ is given by

$$\left[\sqrt{\frac{14 \times 8.44^2}{23.65}}, \sqrt{\frac{14 \times 8.44^2}{6.57}} \right] = [6.49, 12.32].$$

5.2 Parametric Approach to Probability Estimation

Given continuous data x_1, x_2, \dots, x_n , in addition to sample mean \bar{x} and sample standard deviation D , one may also be interested in characterizing the distributional properties including the probability of detection. Say, if the measured data is the time to detect, and there is a certain time limit, say x_0 so that only test data that fall below x_0 is declared a success, then the goal is the estimation of the probability $P(X \leq x)$ or the cumulative distribution function (CDF) at $x = x_0$, i.e. $F(x)$, as defined in Subsection 1.3. So if X is a normal population $N(\mu, \sigma^2)$, then

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

where Φ is the CDF of the standard normal. A sample estimate of $F(x_0)$ is then given by

$$\hat{F}(x_0) = \Phi\left(\frac{x_0 - \bar{x}}{D}\right). \quad (28)$$

Of course, Φ should be replaced by some other function if it is deemed that the underlying data may fit some other parametric probability models. For example, if log-normal model is considered, $\log(X) \sim N(\mu, \sigma^2)$, then

$$\hat{F}(x_0) = \Phi\left(\frac{\log(x_0) - \bar{x}_L}{D_L}\right),$$

where \bar{x}_L and D_L are the sample mean and sample standard deviation on the log-transformed data $(\log(x_1), \dots, \log(x_n))$.

Example 11 Parametric Probability Estimation.

Going back to data in Example 10, Figure 7 shows the parametric fit (red, smooth line) of CDF using (28) with $\bar{x} = 96.8(s)$, $D = 8.44(s)$ for data in Example 10 and it can be compared to the empirical CDF (29) (the black, stair lines). To illustrate how uncertainty can be assigned to the parametric probability estimate, consider a particular scenario when one is only interested in test results where the time to detect is less than 90 seconds, that is, $x_0 = 90(s)$. Then an estimate of the probability of detection $F(x_0)$ is given by $\Phi((90 - 96.8)/8.44) = 0.21$. The 90% confidence interval using the parametric bootstrap method with sample size 1000 is given by $[0.089, 0.358]$ from one simulation exercise. Note that since there are exactly three data points below 90, so if we use methods from Section 2 or Section 4.5, with $n = 15, s = 3$, we can obtain similar point and interval estimates. For example, from Table 9, the credible interval is $[0.090, 0.417]$.

Figure 7 also shows the 90% confidence interval limits at every potential point of interest (green, upper and lower dotted lines) using the same Monte Carlo samples on \bar{x} and D . The median value from the Bootstrap samples is also plotted (green, dotted middle line) and is seen to be close to the parametric fitted line (red, solid smooth line).

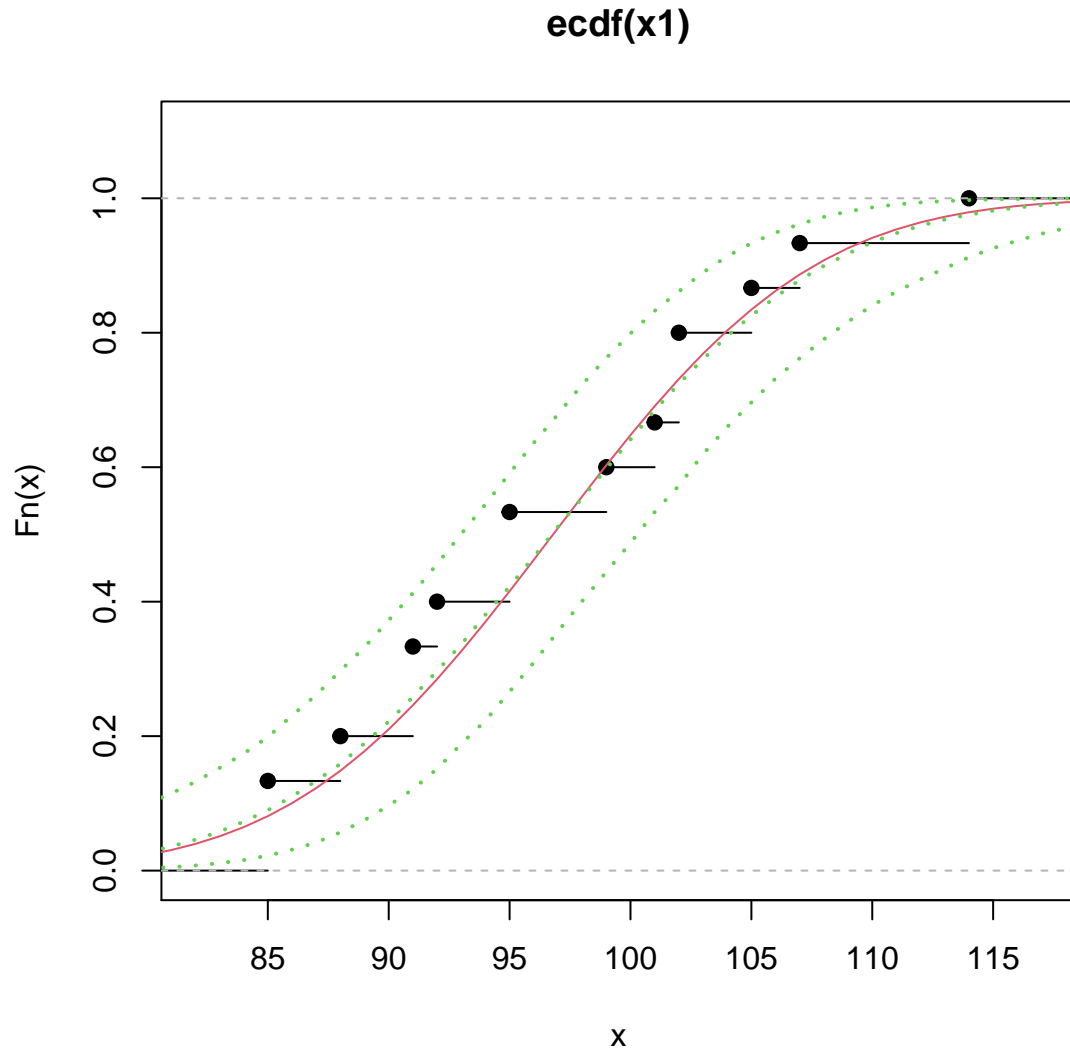


Figure 7: Display of Empirical CDF of Test Data in Example 10 and the Normal Probability Model-based Fit.

Empirical CDF is denoted by the stair-like (broken) lines (in black), and the normal probability fit is denoted by solid smooth line (in red), along with the 90% confidence interval limits from one Monte Carlo simulation run (green, dotted lower and upper lines). The middle dotted green line denotes the median value from the Monte Carlo samples.

This parametric fit can be compared with the empirical CDF function

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{x_i \leq x\}}, \quad (29)$$

which is simply the fraction of number of observed successes $\#\{x_i \leq x\}/n$. Note that the parametric fit (28) is a smooth function of x while $F_n(x)$ is a step function with discontinuity at discrete data points.

To obtain uncertainty estimation for $\hat{F}(x)$ of (28), one can just need to propagate the uncertainty of \bar{x} and D through the function Φ in some way. For example, one approach is through the parametric bootstrap method, or the Monte Carlo method, in which one can generate many realizations of \bar{x} and D through $\bar{x} + D \times N(0, 1)/\sqrt{n}$ and $\sqrt{(n-1)D^2/\chi_{n-1}^2}$ where $N(0, 1)$ and χ_{n-1}^2 are the standard normal and χ^2 random variable with $(n-1)$ degree of freedom, and plug in $\hat{F}(x)$ of (28) to obtain many realizations of it, or Monte Carlo samples. An $100(1-\alpha)\%$ confidence interval for $F(x_0)$ is given the $100\alpha/2\%$ and $100(1-\alpha/2)\%$ sample quantiles of the Monte Carlo $\hat{F}(x_0)$ samples.

5.3 Sample Quantiles and Related Confidence Intervals

Given n measured data X_1, X_2, \dots, X_n which are independent and have the same distribution (cdf) $F(x)$. If X_1 is continuous or $F(x)$ is strictly increasing, the equation

$$F(x) = p, 0 < p < 1,$$

has a unique solution, say $x = \xi_p$, which we call the (population, or theoretical) $100p\%$ th quantile. Thus, $\xi_{\frac{1}{2}}$ is the *median* of the distribution.

When X_1 is discrete, ξ_p is defined by

$$P(X < \xi_p) \leq p \leq P(X \leq \xi_p).$$

This may define ξ_p uniquely or ξ_p lies in an interval. In general, one denote $\xi_p = F^{-1}(x)$, where

$$F^{-1}(x) = \inf\{x : F(x) > p\}.$$

One can define the r th order statistics $X_{(r)}$ as the r th smallest value, or that the order statistics $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ are the sorted values, increasing from smallest to largest among the n measurements X_1, X_2, \dots, X_n . Note that when F is continuous, $F(\xi_p) = p$ and $F(X_{(1)}), F(X_{(2)}), \dots, F(X_{(n)})$ have the same distribution as ordered statistics from the uniform distribution $U(0, 1)$, so it follows that the random interval $[X_{(r)}, X_{(t)}]$ for $r < t$ covers ξ_p with a probability which depends on r, t, n , and p , but not on $F(x)$, thus allowing the construction of distribution-free confidence intervals for ξ_p . It can be shown that

$$\begin{aligned} P(X_{(r)} \leq \xi_p \leq X_{(t)}) &= I_p(r, n-r+1) - I_p(t, n-t+1) \\ &= \sum_{k=r}^{t-1} \binom{n}{k} p^k (1-p)^{n-k}, \end{aligned} \tag{30}$$

Where $I_p(\gamma, \beta)$ is the incomplete Beta function (32), while (30) is simply the Binomial interval probability (11) of Section 3.1. The probability (30) is also known to be still valid

as an lower bound for $P(X_{(r)} \leq \xi_p \leq X_{(t)})$ and upper bound for $P(X_{(r)} < \xi_p < X_{(t)})$ even when F is discrete.

So if we choose r , and t from the Binomial probability distribution $S \sim \text{Bin}(n, p)$ so that (30) is equal to or bigger than $(1 - \alpha)$, for example,

$$P(S < r) < \alpha/2, P(S \geq t) = 1 - P(S \leq t - 1) < \alpha/2, \quad (31)$$

we obtain a confidence interval for ξ_p , given by $[X_{(r)}, X_{(t)}]$, which covers ξ_p with a probability at least $(1 - \alpha)$.

Alternatively, if we choose to approximate (30) by normal approximation as in (13), we can choose

$$\frac{(r - 0.5 - np)}{\sigma} = z_{\alpha/2}, \frac{(t - 1 + 0.5 - np)}{\sigma} = z_{1-\alpha/2},$$

where $\sigma = \sqrt{np(1-p)}$, we get

$$r = np + 0.5 - z_{1-\alpha/2}\sigma,$$

if it is an integer or the integer part of the obtained value, and

$$t = np + 0.5 + z_{1-\alpha/2}\sigma,$$

if it is integer, or the smallest integer that exceeds the obtained.

A special case of quantile is $\xi_{\frac{1}{2}}$, the median. So (30) provides a way to compute confidence intervals for median with $p = \frac{1}{2}$. The quarter quantiles $\xi_{\frac{1}{4}}$ and $\xi_{\frac{3}{4}}$ are also of interest as together they define the interquartile range $\xi_{\frac{3}{4}} - \xi_{\frac{1}{4}}$, an alternative measure of the spread of a distribution.

Point estimation of median from a sample of data X_1, X_2, \dots, X_n is given by the order statistics $X_{((n-1)/2+1)}$ if n is an odd number, or $(X_{(n/2)} + X_{(n/2+1)})/2$ if n is an even number. The estimate of ξ_p in general is defined by the order statistics $X_{([np])}$ where $[np]$ denotes the integer or the integer part of np .

Example 12 Confidence Intervals For Quantiles

Using the setup of Example 10, we have $n = 15$ and the ordered data are:

$$85(1), 85(2), 88(3), 91(4), 91(5), 92(6), 95(7), 95(8), 99(9), 101(10), 102(11), \\ 102(12), 105(13), 107(14), 114(15).$$

and the sample median is given by $X(8)$ which is 95. To find the $1 - \alpha$ confidence interval with $\alpha = 0.10$ for median $\xi_{\frac{1}{2}}$, note that

$$p = \frac{1}{2}, \sigma = \sqrt{15 \times \frac{1}{2} \left(1 - \frac{1}{2}\right)} = 1.94, np = 15 \times \frac{1}{2} = 7.5.$$

So the normal approximation gives $7.5 + 0.5 \pm 1.645 \times 1.93 = 8 \pm 3.17$, which gives an interval $[X_{(4)}, X_{(12)}]$, that is, $[91, 102]$.

Appendix A: Technical Notes and Calculation Details

A Note on Beta Distribution and Calculation

The Beta distribution with parameters $\gamma > 0$ and $\beta > 0$, denoted by $Be(\gamma, \beta)$ has density function

$$b(p; \gamma, \beta) = \frac{p^{\gamma-1}(1-p)^{\beta-1}}{B(\gamma, \beta)}, 0 < p < 1$$

where

$$B(\gamma, \beta) = \int_0^1 p^{\gamma-1}(1-p)^{\beta-1} dp = \Gamma(\gamma)\Gamma(\beta)/\Gamma(\gamma+\beta).$$

The $100(1 - \alpha)\%$ quantile $g_{\alpha; \gamma, \beta}$ of a Beta distribution is defined through the incomplete Beta function:

$$I_g(\gamma, \beta) = \int_0^g \frac{p^{\gamma-1}(1-p)^{\beta-1}}{B(\gamma, \beta)} dp = 1 - \alpha \tag{32}$$

and is available from most standard software packages. For example, in the open source statistics package R [14], which contains a set of functions related to the Beta distribution, including one *qbeta* for computing the quantile of Beta distribution, with inputs including the α for probability, and shape parameters γ, β , all of which can be specified as vectors to make the computation fast when many applications are called for in practice.

There is also an interesting connection between incomplete Beta function and the cumulative distribution function of the binomial distribution [8],

$$\sum_s^n \binom{n}{k} p^k (1-p)^{n-k} = I_p(s, n-s+1), \tag{33}$$

which can be proved by integration by parts and thus

$$\sum_0^s \binom{n}{k} p^k (1-p)^{n-k} = 1 - I_p(s+1, n-s). \tag{34}$$

This justifies the calculation of Clopper-Pearson endpoints in Equation (20).

A Note on the Central Limit Theorem

A simple version of the famous Central Limit Theorem is the following. Let X_1, X_2, \dots, X_n be random variables which are independent and identically distributed with mean $\mu = EX_1$ and $\sigma^2 = \text{var}(X_1) = E(X_1 - \mu)^2$. Then,

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n X_i - \mu \right) \xrightarrow{d} N(0, \sigma^2), \tag{35}$$

where the limit \xrightarrow{d} denotes convergence in distribution. That is, the cumulative distribution function (CDF) of $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ converges to the CDF of $N(\mu, \frac{\sigma^2}{n})$ at every point,

$$P(\sqrt{n}(\bar{X} - \mu)/\sigma \leq z) \rightarrow \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx,$$

for every real number $-\infty < z < \infty$.

Appendix B: Bayesian Intervals Computation and Tables

Computing Bayesian Intervals for Binomial Data

Computing Bayesian interval for binomial proportion involves simply plugging in the numbers for the probability credible level, α , and the shape parameters $\gamma + s, \beta + n - s$ into the quantile function, with the $100(1 - \alpha)\%$ lower limit and upper limit being given by

$$[LL, UL] = [qbeta(\alpha/2, \gamma + s, \beta + n - s), qbeta(1 - \alpha/2, \gamma + s, \beta + n - s)]$$

where γ, β are the parameters in the Beta prior, and n, s are the observed data.

Tables for Bayesian Credible Intervals

For illustration and convenience of practical workers, we provide three tables, Table 6, Table 7, Table 8 which give the 95% limits of the Bayesian likelihood intervals (corresponding to Beta prior $\gamma = 1, \beta = 1$ or the uniform prior) for binomial data with sample size from $n = 7$ through 30.

In addition, we also provide the 90% limits of the Bayesian likelihood intervals (corresponding to uniform prior) for binomial data with sample size from $n = 7$ through 30, in Table 9, Table 10, Table 11, and the 80% limits of the Bayesian likelihood intervals with sample size from $n = 7$ through 30 in Table 12, Table 13, Table 14.

Acknowledgments

The writing of this chapter is under the close guidance of Leticia S. Pibida and Dennis Leber. Their advices on the structure and suggestions on pedagogy and writing for general audience are gratefully appreciated. James Yen pointed out a mistake on the use of continuity correction in an earlier version and provided the data set that is used for all the examples in Section 5. Dennis Leber brought to my attention [5], which has influenced the presentations in Section 1.3 and Section 5.2. Both technical reviewers Charles Hagwood and Ronald Collé provided me with good comments and suggestions which have led to much improved final version.

References

- [1] A. Agresti and B. Coull (1998). Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions. *The American Statistician*, May 1998, Vol.52, No.2, 119-126.
- [2] A. Agresti and B. Caffo (2000). Simple and Effective Confidence Intervals for Proportions and Differences of Proportions Result from Adding Two Successes and Two Failures. *The American Statistician*, November 2000, Vol.54, No.4, 280-288.
- [3] J.O.Berger (1985). *Statistical Decision Theory and Bayesian Analysis*. Second Edition, Springer-Verlag, New York.
- [4] D.A. Berry (1996). *Statistics: A Bayesian Perspective*. Belmont, CA: Wadsworth.
- [5] V.Bram Lillard (2017). Censored Data Analysis for Performance Data, Conference Presentation, Institute for Defense Analyses. Retrievable at <https://testscience.org/wp-content/uploads/sites/10/2017/04/Lillard-Censored-Data-short-2017-Knowledge-Exchange.pdf>.
- [6] L.D. Brown, T.T. Cai and A. DasGupta (2001). Interval Estimation for a Binomial Proportion. *Statistical Science*. Volume 16, Issue 2 (2001), 101-133. doi:10.1214/ss/1009213286. <https://projecteuclid.org/euclid.ss/1009213286>.
- [7] C.J. Clopper and E.S. Pearson (1934). The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika*, Vol.26, No.4, pp.404-413.
- [8] H.O. Hartley and E.R. Fitch (1951). A Chart for the Incomplete Beta-Function and the Cumulative Binomial Distribution. *Biometrika*, Vol.38, No.3/4, pp.423-426.
- [9] W. Hoeffding (1956). On the Distribution of the Number of Successes in Independent Trials, Vol. 27, No. 3, September 1956, *Annals of Mathematical Statistics*, pp.713-721.
- [10] Dennis D. Leber, Leticia S. Pibida, Alexander L. Enders (2019). *Confirming a Performance Threshold with a Binary Experimental Response*. NIST Technical Note (NIST TN) - 2045, <https://doi.org/10.6028/NIST.TN.2045>.
- [11] J. C. Hsu, J. T. G. Hwang, H.-K. Liu, S. J. Ruberg, Confidence Intervals Associated with Tests for Bioequivalence, *Biometrika*, Volume 81, Issue 1, March 1994, Pages 103–114, <https://doi.org/10.1093/biomet/81.1.103>.

- [12] P.M. Lee (1997), *Bayesian Statistics: An Introduction*. Second Edition, 1997. Arnold Publishers, London.
- [13] National Research Council (1998), *Statistics, Testing, and Defense Acquisition: New Approaches and Methodological Improvements*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/6037>.
- [14] R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- [15] S.M. Stigler (1982). Thomas Bayes's Bayesian Inference. *Journal of the Royal Statistical Society Series A (General)*, Vol.145, Part 2, pp.250-258.
- [16] E.B. Wilson (1927). Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association*, Vol.22, 209-212.
- [17] James Yen, Dennis Leber, Leticia Pibida (2020). *Comparing Instruments*. NIST Technical Note (NIST TN). To appear.

s	n=	7	n=	8	n=	9	n=	10	n=	11	n=	12
	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL
0	0.003	0.369	0.003	0.336	0.003	0.308	0.002	0.285	0.002	0.265	0.002	0.247
1	0.032	0.527	0.028	0.482	0.025	0.445	0.023	0.413	0.021	0.385	0.019	0.360
2	0.085	0.651	0.075	0.600	0.067	0.556	0.060	0.518	0.055	0.484	0.050	0.454
3	0.157	0.755	0.137	0.701	0.122	0.652	0.109	0.610	0.099	0.572	0.091	0.538
4	0.245	0.843	0.212	0.788	0.187	0.738	0.167	0.692	0.152	0.651	0.139	0.614
5	0.349	0.915	0.299	0.863	0.262	0.813	0.234	0.766	0.211	0.723	0.192	0.684
6	0.473	0.968	0.400	0.925	0.348	0.878	0.308	0.833	0.277	0.789	0.251	0.749
7	0.631	0.997	0.518	0.972	0.444	0.933	0.390	0.891	0.349	0.848	0.316	0.808
8			0.664	0.997	0.555	0.975	0.482	0.940	0.428	0.901	0.386	0.861
9					0.692	0.997	0.587	0.977	0.516	0.945	0.462	0.909
10							0.715	0.998	0.615	0.979	0.546	0.950
11									0.735	0.998	0.640	0.981
12											0.753	0.998

s	n=	13	n=	14	n=	15	n=	16	n=	17	n=	18
	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL
0	0.002	0.232	0.002	0.218	0.002	0.206	0.001	0.195	0.001	0.185	0.001	0.176
1	0.018	0.339	0.017	0.319	0.016	0.302	0.015	0.287	0.014	0.273	0.013	0.260
2	0.047	0.428	0.043	0.405	0.040	0.383	0.038	0.364	0.036	0.347	0.034	0.331
3	0.084	0.508	0.078	0.481	0.073	0.456	0.068	0.434	0.064	0.414	0.061	0.396
4	0.128	0.581	0.118	0.551	0.110	0.524	0.103	0.499	0.097	0.476	0.091	0.456
5	0.177	0.649	0.163	0.616	0.152	0.587	0.142	0.560	0.133	0.535	0.126	0.512
6	0.230	0.711	0.213	0.677	0.198	0.646	0.184	0.617	0.173	0.590	0.163	0.566
7	0.289	0.770	0.266	0.734	0.247	0.701	0.230	0.671	0.215	0.643	0.203	0.616
8	0.351	0.823	0.323	0.787	0.299	0.753	0.278	0.722	0.260	0.692	0.244	0.665
9	0.419	0.872	0.384	0.837	0.354	0.802	0.329	0.770	0.308	0.740	0.289	0.711
10	0.492	0.916	0.449	0.882	0.413	0.848	0.383	0.816	0.357	0.785	0.335	0.756
11	0.572	0.953	0.519	0.922	0.476	0.890	0.440	0.858	0.410	0.827	0.384	0.797
12	0.661	0.982	0.595	0.957	0.544	0.927	0.501	0.897	0.465	0.867	0.434	0.837
13	0.768	0.998	0.681	0.983	0.617	0.960	0.566	0.932	0.524	0.903	0.488	0.874
14			0.782	0.998	0.698	0.984	0.636	0.962	0.586	0.936	0.544	0.909
15					0.794	0.998	0.713	0.985	0.653	0.964	0.604	0.939
16							0.805	0.999	0.727	0.986	0.669	0.966
17									0.815	0.999	0.740	0.987
18											0.824	0.999

Table 6: 95% Limits $[LL, UL]$ for Bayesian Intervals for Binomial Data with Sample Size $n = 7$ Through 18

s	n=	19	n=	20	n=	21	n=	22	n=	23	n=	24
	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL
0	0.001	0.168	0.001	0.161	0.001	0.154	0.001	0.148	0.001	0.142	0.001	0.137
1	0.012	0.249	0.012	0.238	0.011	0.228	0.011	0.219	0.010	0.211	0.010	0.204
2	0.032	0.317	0.030	0.304	0.029	0.292	0.028	0.280	0.027	0.270	0.025	0.260
3	0.057	0.379	0.054	0.363	0.052	0.349	0.050	0.336	0.047	0.324	0.045	0.312
4	0.087	0.437	0.082	0.419	0.078	0.403	0.075	0.388	0.071	0.374	0.068	0.361
5	0.119	0.491	0.113	0.472	0.107	0.454	0.102	0.437	0.098	0.422	0.094	0.407
6	0.154	0.543	0.146	0.522	0.139	0.502	0.132	0.484	0.126	0.467	0.121	0.451
7	0.191	0.592	0.181	0.570	0.172	0.549	0.164	0.529	0.156	0.511	0.149	0.494
8	0.231	0.639	0.218	0.616	0.207	0.593	0.197	0.573	0.188	0.553	0.180	0.535
9	0.272	0.685	0.257	0.660	0.244	0.636	0.232	0.615	0.221	0.594	0.211	0.575
10	0.315	0.728	0.298	0.702	0.282	0.678	0.268	0.655	0.256	0.634	0.244	0.613
11	0.361	0.769	0.340	0.743	0.322	0.718	0.306	0.694	0.291	0.672	0.278	0.651
12	0.408	0.809	0.384	0.782	0.364	0.756	0.345	0.732	0.328	0.709	0.313	0.687
13	0.457	0.846	0.430	0.819	0.407	0.793	0.385	0.768	0.366	0.744	0.349	0.722
14	0.509	0.881	0.478	0.854	0.451	0.828	0.427	0.803	0.406	0.779	0.387	0.756
15	0.563	0.913	0.528	0.887	0.498	0.861	0.471	0.836	0.447	0.812	0.425	0.789
16	0.621	0.943	0.581	0.918	0.546	0.893	0.516	0.868	0.489	0.844	0.465	0.820
17	0.683	0.968	0.637	0.946	0.597	0.922	0.563	0.898	0.533	0.874	0.506	0.851
18	0.751	0.988	0.696	0.970	0.651	0.948	0.612	0.925	0.578	0.902	0.549	0.879
19	0.832	0.999	0.762	0.988	0.708	0.971	0.664	0.950	0.626	0.929	0.593	0.906
20			0.839	0.999	0.772	0.989	0.720	0.972	0.676	0.953	0.639	0.932
21					0.846	0.999	0.781	0.989	0.730	0.973	0.688	0.955
22							0.852	0.999	0.789	0.990	0.740	0.975
23									0.858	0.999	0.796	0.990
24											0.863	0.999

Table 7: 95% Limits $[LL, UL]$ for Bayesian Intervals for Binomial Data with Sample Size $n = 19$ Through 24

s	n=	25	n=	26	n=	27	n=	28	n=	29	n=	30
	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL
0	0.001	0.132	0.001	0.128	0.001	0.123	0.001	0.119	0.001	0.116	0.001	0.112
1	0.009	0.196	0.009	0.190	0.009	0.183	0.008	0.178	0.008	0.172	0.008	0.167
2	0.024	0.251	0.024	0.243	0.023	0.235	0.022	0.228	0.021	0.221	0.020	0.214
3	0.044	0.302	0.042	0.292	0.040	0.282	0.039	0.274	0.038	0.265	0.036	0.258
4	0.066	0.349	0.063	0.337	0.061	0.327	0.058	0.317	0.056	0.307	0.055	0.298
5	0.090	0.394	0.086	0.381	0.083	0.369	0.080	0.358	0.077	0.347	0.075	0.337
6	0.116	0.436	0.111	0.423	0.107	0.410	0.103	0.397	0.099	0.386	0.096	0.375
7	0.143	0.478	0.138	0.463	0.132	0.449	0.127	0.435	0.123	0.423	0.119	0.411
8	0.172	0.518	0.165	0.502	0.159	0.487	0.153	0.472	0.147	0.459	0.142	0.446
9	0.202	0.557	0.194	0.540	0.186	0.524	0.179	0.508	0.173	0.494	0.167	0.480
10	0.234	0.594	0.224	0.576	0.215	0.559	0.207	0.543	0.199	0.528	0.192	0.514
11	0.266	0.631	0.255	0.612	0.245	0.594	0.235	0.577	0.227	0.561	0.218	0.546
12	0.299	0.666	0.287	0.647	0.275	0.628	0.264	0.611	0.255	0.594	0.245	0.578
13	0.334	0.701	0.319	0.681	0.306	0.661	0.294	0.643	0.283	0.626	0.273	0.609
14	0.369	0.734	0.353	0.713	0.339	0.694	0.325	0.675	0.313	0.657	0.302	0.640
15	0.406	0.766	0.388	0.745	0.372	0.725	0.357	0.706	0.343	0.687	0.331	0.669
16	0.443	0.798	0.424	0.776	0.406	0.755	0.389	0.736	0.374	0.717	0.360	0.698
17	0.482	0.828	0.460	0.806	0.441	0.785	0.423	0.765	0.406	0.745	0.391	0.727
18	0.522	0.857	0.498	0.835	0.476	0.814	0.457	0.793	0.439	0.773	0.422	0.755
19	0.564	0.884	0.537	0.862	0.513	0.841	0.492	0.821	0.472	0.801	0.454	0.782
20	0.606	0.910	0.577	0.889	0.551	0.868	0.528	0.847	0.506	0.827	0.486	0.808
21	0.651	0.934	0.619	0.914	0.590	0.893	0.565	0.873	0.541	0.853	0.520	0.833
22	0.698	0.956	0.663	0.937	0.631	0.917	0.603	0.897	0.577	0.877	0.554	0.858
23	0.749	0.976	0.708	0.958	0.673	0.939	0.642	0.920	0.614	0.901	0.589	0.881
24	0.804	0.991	0.757	0.976	0.718	0.960	0.683	0.942	0.653	0.923	0.625	0.904
25	0.868	0.999	0.810	0.991	0.765	0.977	0.726	0.961	0.693	0.944	0.663	0.925
26			0.872	0.999	0.817	0.991	0.772	0.978	0.735	0.962	0.702	0.945
27					0.877	0.999	0.822	0.992	0.779	0.979	0.742	0.964
28							0.881	0.999	0.828	0.992	0.786	0.980
29									0.884	0.999	0.833	0.992
30											0.888	0.999

Table 8: 95% Limits $[LL, UL]$ for Bayesian Intervals for Binomial Data with Sample Size $n = 25$ Through 30

s	n= 7		n= 8		n= 9		n= 10		n= 11		n= 12	
0	0.006	0.312	0.006	0.283	0.005	0.259	0.005	0.238	0.004	0.221	0.004	0.206
1	0.046	0.471	0.041	0.429	0.037	0.394	0.033	0.364	0.030	0.339	0.028	0.316
2	0.111	0.600	0.098	0.550	0.087	0.507	0.079	0.470	0.072	0.438	0.066	0.410
3	0.193	0.711	0.169	0.655	0.150	0.607	0.135	0.564	0.123	0.527	0.113	0.495
4	0.289	0.807	0.251	0.749	0.222	0.696	0.200	0.650	0.181	0.609	0.166	0.573
5	0.400	0.889	0.345	0.831	0.304	0.778	0.271	0.729	0.245	0.685	0.224	0.645
6	0.529	0.954	0.450	0.902	0.393	0.850	0.350	0.800	0.315	0.755	0.287	0.713
7	0.688	0.994	0.571	0.959	0.493	0.913	0.436	0.865	0.391	0.819	0.355	0.776
8			0.717	0.994	0.606	0.963	0.530	0.921	0.473	0.877	0.427	0.834
9					0.741	0.995	0.636	0.967	0.562	0.928	0.505	0.887
10							0.762	0.995	0.661	0.970	0.590	0.934
11									0.779	0.996	0.684	0.972
12											0.794	0.996
s	n= 13		n= 14		n= 15		n= 16		n= 17		n= 18	
0	0.004	0.193	0.003	0.181	0.003	0.171	0.003	0.162	0.003	0.153	0.003	0.146
1	0.026	0.297	0.024	0.279	0.023	0.264	0.021	0.250	0.020	0.238	0.019	0.226
2	0.061	0.385	0.057	0.363	0.053	0.344	0.050	0.326	0.047	0.310	0.044	0.296
3	0.104	0.466	0.097	0.440	0.090	0.417	0.085	0.396	0.080	0.377	0.075	0.359
4	0.153	0.540	0.142	0.511	0.132	0.484	0.124	0.461	0.116	0.439	0.110	0.419
5	0.206	0.610	0.191	0.577	0.178	0.548	0.166	0.522	0.156	0.498	0.147	0.476
6	0.264	0.675	0.244	0.640	0.227	0.609	0.212	0.580	0.199	0.554	0.188	0.530
7	0.325	0.736	0.300	0.700	0.279	0.667	0.260	0.636	0.244	0.608	0.230	0.582
8	0.390	0.794	0.360	0.756	0.333	0.721	0.311	0.689	0.291	0.659	0.274	0.632
9	0.460	0.847	0.423	0.809	0.391	0.773	0.364	0.740	0.341	0.709	0.320	0.680
10	0.534	0.896	0.489	0.858	0.452	0.822	0.420	0.788	0.392	0.756	0.368	0.726
11	0.615	0.939	0.560	0.903	0.516	0.868	0.478	0.834	0.446	0.801	0.418	0.770
12	0.703	0.974	0.637	0.943	0.583	0.910	0.539	0.876	0.502	0.844	0.470	0.812
13	0.807	0.996	0.721	0.976	0.656	0.947	0.604	0.915	0.561	0.884	0.524	0.853
14			0.819	0.997	0.736	0.977	0.674	0.950	0.623	0.920	0.581	0.890
15					0.829	0.997	0.750	0.979	0.690	0.953	0.641	0.925
16							0.838	0.997	0.762	0.980	0.704	0.956
17									0.847	0.997	0.774	0.981
18											0.854	0.997

Table 9: 90% Limits $[LL, UL]$ for Bayesian Intervals with Sample Size $n=7$ Through 18.

s	n= 19		n= 20		n= 21		n= 22		n= 23		n= 24	
0	0.003	0.139	0.002	0.133	0.002	0.127	0.002	0.122	0.002	0.117	0.002	0.113
1	0.018	0.216	0.017	0.207	0.016	0.198	0.016	0.190	0.015	0.183	0.014	0.176
2	0.042	0.283	0.040	0.271	0.038	0.259	0.037	0.249	0.035	0.240	0.034	0.231
3	0.071	0.344	0.068	0.329	0.065	0.316	0.062	0.304	0.059	0.292	0.057	0.282
4	0.104	0.401	0.099	0.384	0.094	0.369	0.090	0.355	0.086	0.342	0.082	0.330
5	0.140	0.456	0.132	0.437	0.126	0.420	0.120	0.404	0.115	0.389	0.110	0.375
6	0.177	0.508	0.168	0.487	0.160	0.468	0.152	0.451	0.146	0.435	0.139	0.420
7	0.217	0.558	0.206	0.536	0.196	0.515	0.186	0.496	0.178	0.479	0.170	0.462
8	0.259	0.606	0.245	0.583	0.233	0.561	0.222	0.540	0.212	0.521	0.202	0.504
9	0.302	0.653	0.286	0.628	0.271	0.605	0.258	0.583	0.246	0.563	0.236	0.544
10	0.347	0.698	0.328	0.672	0.311	0.647	0.296	0.625	0.282	0.603	0.270	0.583
11	0.394	0.741	0.372	0.714	0.353	0.689	0.335	0.665	0.319	0.642	0.305	0.621
12	0.442	0.783	0.417	0.755	0.395	0.729	0.375	0.704	0.358	0.681	0.341	0.659
13	0.492	0.823	0.464	0.794	0.439	0.767	0.417	0.742	0.397	0.718	0.379	0.695
14	0.544	0.860	0.513	0.832	0.485	0.804	0.460	0.778	0.437	0.754	0.417	0.730
15	0.599	0.896	0.563	0.868	0.532	0.840	0.504	0.814	0.479	0.788	0.456	0.764
16	0.656	0.929	0.616	0.901	0.580	0.874	0.549	0.848	0.521	0.822	0.496	0.798
17	0.717	0.958	0.671	0.932	0.631	0.906	0.596	0.880	0.565	0.854	0.538	0.830
18	0.784	0.982	0.729	0.960	0.684	0.935	0.645	0.910	0.611	0.885	0.580	0.861
19	0.861	0.997	0.793	0.983	0.741	0.962	0.696	0.938	0.658	0.914	0.625	0.890
20			0.867	0.998	0.802	0.984	0.751	0.963	0.708	0.941	0.670	0.918
21					0.873	0.998	0.810	0.984	0.760	0.965	0.718	0.943
22							0.878	0.998	0.817	0.985	0.769	0.966
23									0.883	0.998	0.824	0.986
24											0.887	0.998

Table 10: 90% Limits $[LL, UL]$ for Bayesian Intervals with Sample Size $n=19$ Through 24.

s	n= 25		n= 26		n= 27		n= 28		n= 29		n= 30	
0	0.002	0.109	0.002	0.105	0.002	0.101	0.002	0.098	0.002	0.095	0.002	0.092
1	0.014	0.170	0.013	0.164	0.013	0.159	0.012	0.153	0.012	0.149	0.012	0.144
2	0.032	0.223	0.031	0.215	0.030	0.208	0.029	0.202	0.028	0.195	0.027	0.189
3	0.054	0.272	0.052	0.263	0.050	0.254	0.049	0.246	0.047	0.239	0.045	0.232
4	0.079	0.318	0.076	0.308	0.073	0.298	0.070	0.288	0.068	0.280	0.066	0.271
5	0.106	0.363	0.101	0.351	0.098	0.339	0.094	0.329	0.091	0.319	0.088	0.310
6	0.134	0.405	0.129	0.392	0.124	0.380	0.119	0.368	0.115	0.357	0.111	0.347
7	0.163	0.447	0.157	0.432	0.151	0.419	0.145	0.406	0.140	0.394	0.135	0.383
8	0.194	0.487	0.186	0.471	0.179	0.457	0.172	0.443	0.166	0.430	0.161	0.418
9	0.226	0.526	0.217	0.509	0.208	0.494	0.200	0.479	0.193	0.465	0.187	0.452
10	0.258	0.564	0.248	0.547	0.238	0.530	0.229	0.514	0.221	0.499	0.213	0.485
11	0.292	0.602	0.280	0.583	0.269	0.565	0.259	0.549	0.250	0.533	0.241	0.518
12	0.327	0.638	0.313	0.618	0.301	0.600	0.289	0.583	0.279	0.566	0.269	0.550
13	0.362	0.673	0.347	0.653	0.333	0.634	0.320	0.616	0.308	0.598	0.297	0.582
14	0.398	0.708	0.382	0.687	0.366	0.667	0.352	0.648	0.339	0.630	0.327	0.613
15	0.436	0.742	0.417	0.720	0.400	0.699	0.384	0.680	0.370	0.661	0.357	0.643
16	0.474	0.774	0.453	0.752	0.435	0.731	0.417	0.711	0.402	0.692	0.387	0.673
17	0.513	0.806	0.491	0.783	0.470	0.762	0.451	0.741	0.434	0.721	0.418	0.703
18	0.553	0.837	0.529	0.814	0.506	0.792	0.486	0.771	0.467	0.750	0.450	0.731
19	0.595	0.866	0.568	0.843	0.543	0.821	0.521	0.800	0.501	0.779	0.482	0.759
20	0.637	0.894	0.608	0.871	0.581	0.849	0.557	0.828	0.535	0.807	0.515	0.787
21	0.682	0.921	0.649	0.899	0.620	0.876	0.594	0.855	0.570	0.834	0.548	0.813
22	0.728	0.946	0.692	0.924	0.661	0.902	0.632	0.881	0.606	0.860	0.582	0.839
23	0.777	0.968	0.737	0.948	0.702	0.927	0.671	0.906	0.643	0.885	0.617	0.865
24	0.830	0.986	0.785	0.969	0.746	0.950	0.712	0.930	0.681	0.909	0.653	0.889
25	0.891	0.998	0.836	0.987	0.792	0.970	0.754	0.951	0.720	0.932	0.690	0.912
26			0.895	0.998	0.841	0.987	0.798	0.971	0.761	0.953	0.729	0.934
27					0.899	0.998	0.847	0.988	0.805	0.972	0.768	0.955
28							0.902	0.998	0.851	0.988	0.811	0.973
29									0.905	0.998	0.856	0.988
30											0.908	0.998

Table 11: 90% Limits $[LL, UL]$ for Bayesian Intervals with Sample Size $n=25$ Through 30.

s	n= 7		n= 8		n= 9		n= 10		n= 11		n= 12	
0	0.013	0.250	0.012	0.226	0.010	0.206	0.010	0.189	0.009	0.175	0.008	0.162
1	0.069	0.406	0.061	0.368	0.055	0.337	0.049	0.310	0.045	0.287	0.042	0.268
2	0.147	0.538	0.129	0.490	0.116	0.450	0.105	0.415	0.096	0.386	0.088	0.360
3	0.240	0.655	0.210	0.599	0.188	0.552	0.169	0.511	0.154	0.475	0.142	0.444
4	0.345	0.760	0.301	0.699	0.267	0.646	0.241	0.599	0.219	0.559	0.201	0.523
5	0.462	0.853	0.401	0.790	0.354	0.733	0.318	0.682	0.288	0.638	0.264	0.598
6	0.594	0.931	0.510	0.871	0.448	0.812	0.401	0.759	0.362	0.712	0.331	0.669
7	0.750	0.987	0.632	0.939	0.550	0.884	0.489	0.831	0.441	0.781	0.402	0.736
8			0.774	0.988	0.663	0.945	0.585	0.895	0.525	0.846	0.477	0.799
9					0.794	0.990	0.690	0.951	0.614	0.904	0.556	0.858
10							0.811	0.990	0.713	0.955	0.640	0.912
11									0.825	0.991	0.732	0.958
12											0.838	0.992
s	n= 13		n= 14		n= 15		n= 16		n= 17		n= 18	
0	0.007	0.152	0.007	0.142	0.007	0.134	0.006	0.127	0.006	0.120	0.006	0.114
1	0.039	0.251	0.036	0.236	0.034	0.222	0.032	0.210	0.030	0.199	0.028	0.190
2	0.081	0.337	0.076	0.317	0.071	0.300	0.067	0.284	0.063	0.269	0.059	0.257
3	0.131	0.417	0.122	0.393	0.114	0.371	0.107	0.352	0.101	0.334	0.095	0.319
4	0.185	0.492	0.172	0.464	0.161	0.439	0.151	0.416	0.142	0.396	0.134	0.378
5	0.243	0.563	0.226	0.532	0.210	0.504	0.197	0.478	0.185	0.455	0.175	0.434
6	0.305	0.631	0.282	0.596	0.263	0.565	0.246	0.537	0.231	0.512	0.218	0.489
7	0.369	0.695	0.342	0.658	0.318	0.625	0.297	0.594	0.279	0.567	0.263	0.541
8	0.437	0.757	0.404	0.718	0.375	0.682	0.350	0.650	0.329	0.620	0.310	0.592
9	0.508	0.815	0.468	0.774	0.435	0.737	0.406	0.703	0.380	0.671	0.358	0.642
10	0.583	0.869	0.536	0.828	0.496	0.790	0.463	0.754	0.433	0.721	0.408	0.690
11	0.663	0.919	0.607	0.878	0.561	0.839	0.522	0.803	0.488	0.769	0.459	0.737
12	0.749	0.961	0.683	0.924	0.629	0.886	0.584	0.849	0.545	0.815	0.511	0.782
13	0.848	0.993	0.764	0.964	0.700	0.929	0.648	0.893	0.604	0.858	0.566	0.825
14			0.858	0.993	0.778	0.966	0.716	0.933	0.666	0.899	0.622	0.866
15					0.866	0.993	0.790	0.968	0.731	0.937	0.681	0.905
16							0.873	0.994	0.801	0.970	0.743	0.941
17									0.880	0.994	0.810	0.972
18											0.886	0.994

Table 12: 80% Equal-tailed Bayesian Intervals For $n = 7$ Through 18

s	n= 19		n= 20		n= 21		n= 22		n= 23		n= 24	
0	0.005	0.109	0.005	0.104	0.005	0.099	0.005	0.095	0.004	0.091	0.004	0.088
1	0.027	0.181	0.026	0.173	0.024	0.166	0.023	0.159	0.022	0.153	0.021	0.147
2	0.056	0.245	0.054	0.234	0.051	0.224	0.049	0.215	0.047	0.207	0.045	0.199
3	0.090	0.304	0.086	0.291	0.082	0.279	0.078	0.268	0.075	0.258	0.072	0.248
4	0.127	0.361	0.121	0.345	0.115	0.331	0.110	0.318	0.105	0.306	0.101	0.295
5	0.166	0.415	0.158	0.397	0.150	0.381	0.143	0.366	0.137	0.352	0.131	0.340
6	0.207	0.467	0.196	0.448	0.187	0.430	0.178	0.413	0.170	0.398	0.163	0.383
7	0.249	0.518	0.236	0.497	0.225	0.477	0.214	0.459	0.205	0.442	0.196	0.426
8	0.293	0.567	0.278	0.544	0.264	0.523	0.252	0.503	0.241	0.484	0.230	0.467
9	0.338	0.615	0.321	0.590	0.305	0.568	0.290	0.546	0.277	0.526	0.265	0.508
10	0.385	0.662	0.364	0.636	0.346	0.611	0.330	0.589	0.315	0.567	0.301	0.548
11	0.433	0.707	0.410	0.679	0.389	0.654	0.370	0.630	0.353	0.608	0.338	0.587
12	0.482	0.751	0.456	0.722	0.432	0.695	0.411	0.670	0.392	0.647	0.375	0.625
13	0.533	0.793	0.503	0.764	0.477	0.736	0.454	0.710	0.433	0.685	0.413	0.662
14	0.585	0.834	0.552	0.804	0.523	0.775	0.497	0.748	0.474	0.723	0.452	0.699
15	0.639	0.873	0.603	0.842	0.570	0.813	0.541	0.786	0.516	0.759	0.492	0.735
16	0.696	0.910	0.655	0.879	0.619	0.850	0.587	0.822	0.558	0.795	0.533	0.770
17	0.755	0.944	0.709	0.914	0.669	0.885	0.634	0.857	0.602	0.830	0.574	0.804
18	0.819	0.973	0.766	0.946	0.721	0.918	0.682	0.890	0.648	0.863	0.617	0.837
19	0.891	0.995	0.827	0.974	0.776	0.949	0.732	0.922	0.694	0.895	0.660	0.869
20			0.896	0.995	0.834	0.976	0.785	0.951	0.742	0.925	0.705	0.899
21					0.901	0.995	0.841	0.977	0.793	0.953	0.752	0.928
22							0.905	0.995	0.847	0.978	0.801	0.955
23									0.909	0.996	0.853	0.979
24											0.912	0.996

Table 13: 80% Equal-tailed Bayesian Intervals For $n = 19$ Through 24

s	n= 25		n= 26		n= 27		n= 28		n= 29		n= 30	
0	0.004	0.085	0.004	0.082	0.004	0.079	0.004	0.076	0.004	0.074	0.003	0.072
1	0.021	0.142	0.020	0.137	0.019	0.132	0.018	0.128	0.018	0.124	0.017	0.120
2	0.043	0.192	0.042	0.185	0.040	0.179	0.039	0.173	0.037	0.168	0.036	0.163
3	0.069	0.239	0.066	0.231	0.064	0.223	0.062	0.216	0.059	0.209	0.057	0.203
4	0.097	0.284	0.093	0.275	0.089	0.265	0.086	0.257	0.083	0.249	0.081	0.241
5	0.126	0.328	0.121	0.317	0.117	0.306	0.112	0.297	0.109	0.287	0.105	0.279
6	0.157	0.370	0.151	0.358	0.145	0.346	0.140	0.335	0.135	0.325	0.130	0.315
7	0.188	0.411	0.181	0.397	0.174	0.385	0.168	0.372	0.162	0.361	0.157	0.350
8	0.221	0.451	0.212	0.436	0.204	0.422	0.197	0.409	0.190	0.397	0.183	0.385
9	0.254	0.491	0.244	0.475	0.235	0.459	0.226	0.445	0.218	0.432	0.211	0.419
10	0.289	0.529	0.277	0.512	0.267	0.496	0.257	0.481	0.248	0.466	0.239	0.453
11	0.324	0.567	0.311	0.549	0.299	0.532	0.288	0.515	0.277	0.500	0.268	0.486
12	0.359	0.604	0.345	0.585	0.331	0.567	0.319	0.550	0.308	0.533	0.297	0.518
13	0.396	0.641	0.380	0.620	0.365	0.601	0.351	0.583	0.338	0.566	0.327	0.550
14	0.433	0.676	0.415	0.655	0.399	0.635	0.384	0.616	0.370	0.599	0.357	0.582
15	0.471	0.711	0.451	0.689	0.433	0.669	0.417	0.649	0.401	0.630	0.387	0.613
16	0.509	0.746	0.488	0.723	0.468	0.701	0.450	0.681	0.434	0.662	0.418	0.643
17	0.549	0.779	0.525	0.756	0.504	0.733	0.485	0.712	0.467	0.692	0.450	0.673
18	0.589	0.812	0.564	0.788	0.541	0.765	0.519	0.743	0.500	0.723	0.482	0.703
19	0.630	0.843	0.603	0.819	0.578	0.796	0.555	0.774	0.534	0.752	0.514	0.732
20	0.672	0.874	0.642	0.849	0.615	0.826	0.591	0.803	0.568	0.782	0.547	0.761
21	0.716	0.903	0.683	0.879	0.654	0.855	0.628	0.832	0.603	0.810	0.581	0.789
22	0.761	0.931	0.725	0.907	0.694	0.883	0.665	0.860	0.639	0.838	0.615	0.817
23	0.808	0.957	0.769	0.934	0.735	0.911	0.703	0.888	0.675	0.865	0.650	0.843
24	0.858	0.979	0.815	0.958	0.777	0.936	0.743	0.914	0.713	0.891	0.685	0.870
25	0.915	0.996	0.863	0.980	0.821	0.960	0.784	0.938	0.751	0.917	0.721	0.895
26			0.918	0.996	0.868	0.981	0.827	0.961	0.791	0.941	0.759	0.919
27					0.921	0.996	0.872	0.982	0.832	0.963	0.797	0.943
28							0.924	0.996	0.876	0.982	0.837	0.964
29									0.926	0.996	0.880	0.983
30											0.928	0.997

Table 14: 80% Equal-tailed Bayesian Intervals For $n = 25$ Through 30