International Journal of Statistics and Probability; Vol. 12, No. 2; March 2023
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Joint Estimation of Binomial Proportions

Kent Riggs[1] & Stephanie Weatherford

[1] Department of Mathematics and Statistics, Stephen F. Austin State University, Nacogdoches, TX 75962, USA

Correspondence: Kent Riggs, PO Box 13040, SFA Station, Nacogdoches, TX 75962, USA

Received: June 1, 2022 Accepted: October 30, 2023 Online Published: February 21, 2023

doi:10.5539/ijsp.v12n2p8 URL: https://doi.org/10.5539/ijsp.v12n2p8

Abstract

Interval estimation of a binomial proportion has had a consistent presence in the statistical literature through the years. Many interval procedures have been developed for a single proportion as well as for the difference of two proportions. However, little work has been conducted on the joint estimation of two binomial proportions. In this paper, we construct four confidence regions for two binomial proportions based on three statistics: the Wald (W), adjusted Wald (W*), score (S), and likelihood ratio (LR) statistics. Once the regions have been established, we compare their coverage probabilities and average areas for different parameter and sample size configurations. For small-to-moderate sample sizes, this paper finds that the three regions based on the W*, S, and LR statistics have good coverage properties, with the score region usually having the smallest average area. Finally, we apply these four confidence regions to some real data in veterinary science and medicine for the joint estimation of important proportions.

1. Introduction

Interval estimation of a binomial proportion has a long and popular history dating back to the development of confidence intervals over eighty years ago (Schilling and Doi 2014). Most confidence intervals are based on large-sample theory and yield very similar results in the case of a large sample, but in the presence of little data, discrepancies occur between the intervals in terms of average width and probability coverage problems. Agresti and Coull (1998) provide a nice overview of the problem. They also consider the common problem of estimating the difference between two proportions. In addition, Viana (1991) considers joint estimation of several binomial proportions from the Bayesian perspective, though not credible regions for such proportions.

While much work has been done on interval estimation of a single proportion (Newcombe 1998a) and the difference of two proportions (Newcombe 1998b), little work has been done on joint estimation of two binomial proportions using a confidence region. In order to motivate the need for such a confidence region, consider the following example. Suppose that $p_1 = .9$ and $p_2 = .8$ for two populations while $p_1 = .2$ and $p_2 = .1$ for two other populations. Interval estimation techniques for the difference of the two proportions would lead to similar results; however, information regarding where the two proportions might lie individually is lost. A joint confidence region for p_1 and p_2 not only allows one to compare the two proportions but also retains information about plausible values of the individual proportions, as well as any interaction between the two proportions on the unit square. Such a confidence region may also be used to test the hypothesis $H_0 : p_1 = p_2 + d$ by checking to see if the line $p_1 = p_2 + d$ lies in the confidence region. While Elliot and Riggs (2015) considered joint estimation of two negative binomial proportions, the only scholarly work that we have found for two binomial proportions is from Reiczigel, Abonyi, and Singer (2008), who proposed an exact confidence region for two binomial proportions. Although this method maintains appropriate coverage properties, it is computationally intensive and requires the application of a complicated algorithm, which may be out of reach for the researcher who is not well versed in computer programming or cannot obtain an appropriate computer application from the author. Another possible joint estimation technique is the rectangular confidence set created by computing two $\sqrt{1-\alpha}$ confidence intervals, one for p_1 and one for p_2. While this method is easy to implement, it results in larger than necessary areas. In this paper, we wish to develop and study more straightforwardly-developed confidence regions that can be used to jointly estimate p_1 and p_2 from two independent binomial distributions.

The remaining sections will be outlined as follows. In Section 2, we propose three confidence regions for p_1 and p_2 by inverting the Wald, score, and likelihood ratio statistics. In Sections 3 and 4, we compare the regions' average areas and coverage probabilities using a Monte Carlo simulation. In Section 5, we apply our regions to two real-world data sets, and finally, in Section 6, we make some remarks regarding our conclusions.

2. Confidence Regions

Suppose we have two populations ($i = 1,2$) for which the presence of a particular attribute is of interest, such as whether or not a person smokes. Let $0 < p_i < 1$ be the proportion of cases in population *i* possessing the attribute. Now, for each population, suppose we gather a random sample of size n_i and that the two samples are independent of each other. Let Y_i be the number of successes from each sample. Given this sampling structure, the joint probability mass function for (Y_1, Y_2) is

$$f(y_1, y_2) = \binom{n_1}{y_1}\binom{n_2}{y_2} p_1^{y_1} p_2^{y_2} (1-p_1)^{n_1-y_1} (1-p_2)^{n_2-y_2} \tag{1}$$

where $y_i = 0,1,...,n_i$ and $0 < p_i < 1$.

Therefore, the log-likelihood function is

$$\ell(p_1, p_2) = C + x_1 \ln(p_1) + (n_1 - y_1)\ln(1-p_1) + x_2 \ln(p_2) + (n_2 - y_2)\ln(1-p_2)$$

where *C* is a constant. From this expression, the maximum likelihood estimator (MLE) of p_i is straightforwardly show to be

$$\hat{p}_i = \frac{x_i}{n_i}.$$

Also, from the log-likelihood function, we find that the diagonal elements of Fisher's information matrix are given by

$$I_{ii}(p_i) = \frac{n_i}{p_i(1-p_i)},$$

while the off-diagonal elements are 0 by virtue of independent samples. Note that $I_{ii}^{-1}(p_i)$ is the variance of $\hat{p}_i$, or that $\sigma_{\hat{p}_i}^2 = \frac{p_i(1-p_i)}{n_i}$.

We propose creating the first confidence region by inverting a Wald statistic. The Wald statistic is

$$W = \frac{(\hat{p}_1 - p_1)^2}{\hat{\sigma}_{\hat{p}_1}^2} + \frac{(\hat{p}_2 - p_2)^2}{\hat{\sigma}_{\hat{p}_2}^2}.$$

Note that *W* is approximately distributed chi-squared with two degrees of freedom for large sample sizes. Therefore, an approximate 100(1-α)% confidence region for p_1 and p_2 is generated by solving

$$W \leq \chi_{1-\alpha}^2(2) \tag{2}$$

where $\chi_{1-\alpha}^2(2)$ is the (1-α)th percentile of a chi-squared distribution with two degrees of freedom. We note that the boundary of the region in (2) is an ellipse that is centered at $(\hat{p}_1, \hat{p}_2)$ with axes lengths of $a_i = 2\sqrt{\chi_{1-\alpha}^2(2)\,\hat{\sigma}_{\hat{p}_i}^2}$. Hence, the area of (2) is given by $4\pi\chi_{1-\alpha}^2(2)\hat{\sigma}_{\hat{p}_1}\hat{\sigma}_{\hat{p}_2}$ and intuitively reduces as either standard error of $\hat{p}_i$ reduces.

Agresti and Coull (1998) propose an "adjusted" Wald confidence interval for a single proportion simply by adding two successes and two failures in an effort to improve probability coverage when the sample size is not large. In a similar approach, we propose an "adjusted" Wald confidence region by inverting the following statistic

$$W^* = \frac{(\tilde{p}_1 - p_1)^2}{\tilde{\sigma}_{\tilde{p}_1}^2} + \frac{(\tilde{p}_2 - p_2)^2}{\tilde{\sigma}_{\tilde{p}_2}^2},$$

where $\tilde{p}_i = \frac{(y_i+2)}{(n_i+4)}$ and $\hat{\sigma}^2_{\tilde{p}_i} = \frac{\tilde{p}_i(1-\tilde{p}_i)}{(n_i+4)}$. Therefore, an approximate 100(1-α)% confidence region for p_1 and p_2 is generated by solving

$$W^* \le \chi^2_{1-\alpha}(2). \tag{3}$$

Similar to (2), the boundary of the region in (3) is an ellipse that is centered at $(\tilde{p}_1, \tilde{p}_2)$ with axes lengths of $b_i = 2\sqrt{\chi^2_{1-\alpha}(2)\,\hat{\sigma}^2_{\tilde{p}_i}}$.

We propose creating the next confidence region by inverting the following score statistic,

$$S = \frac{(y_1 - n_1p_1)^2}{n_1p_1(1-p_1)} + \frac{(y_2 - n_2p_2)^2}{n_2p_2(1-p_2)},$$

which is approximately distributed chi-squared with two degrees of freedom for large sample sizes. Thus, an approximate 100(1-α)% confidence region for p_1 and p_2 is generated by solving

$$S \le \chi^2_{1-\alpha}(2). \tag{4}$$

We propose creating the last confidence region by inverting a likelihood ratio statistic, which is

$$LR = -2[\ell(p_1, p_2) - \ell(\hat{p}_1, \hat{p}_2)],$$

and is approximately distributed chi-squared with two degrees of freedom for large sample sizes. Thus, an approximate 100(1-α)% confidence region for p_1 and p_2 is generated by solving

$$LR \le \chi^2_{1-\alpha}(2). \tag{5}$$

We note that the elliptical properties of (2) and (3) make it easy to graph, while (4) and (5) must be graphed with the aid of a computer.

3. Coverage Probabilities and Average Areas

Given that the confidence regions (2) – (5) are based on asymptotic approximations, the actual coverage probabilities may be substantially different from the nominal coverage probabilities. For each of the four confidence regions, the actual coverage probability is given by

$$CP = \sum_{y_2}\sum_{y_1} \mathbf{I}_{CR}(p_1, p_2)\, f(y_1, y_2),$$

where $\mathbf{I}_{CR}(p_1, p_2)$ is an indicator function that equals one if the pair (p_1, p_2) lies in the region *CR* and equals zero otherwise. *CP* for each confidence region is estimated using a Monte Carlo technique. In particular, for a fixed (p_1, p_2) and (n_1, n_2), (y_1, y_2) is randomly generated from (1). Then, we check if the inequalities in (2) - (5) are satisfied. This process is carried out 10,000 times, and the relative frequency of times the inequalities are satisfied is a suitable estimate of CP for each confidence region.

In addition to comparing the regions' actual probability coverages, we desire to compare their average areas. The area of a region is a measure of the precision of a particular estimation technique. For each of the four confidence regions, the average (expected) area is given by

$$EA = \sum_{y_2}\sum_{y_1} \left(\int_{CR} 1 dp_1 dp_2\right) f(y_1, y_2).$$

Here, $\int_{CR} 1 dp_1 dp_2$ is the area of a single confidence region for fixed value (y_1, y_2). For the Wald region, $\int_{CR} 1 dp_1 dp_2$ $= 4\pi\chi^2_{1-\alpha}(2)\hat{\sigma}_{\hat{p}_1}\hat{\sigma}_{\hat{p}_2}$ because, as explained before, the boundary of (2) forms an ellipse, which has a closed-form representation for its area. Similarly, an explicit expression for $\int_{CR} 1 dp_1 dp_2$ may be obtained for the adjusted Wald region. However, for the score and likelihood regions, $\int_{CR} 1 dp_1 dp_2$ must be approximated because their boundaries do not form ellipses. For both (4) and (5), $\int_{CR} 1 dp_1 dp_2$ is approximated using a Monte Carlo integration technique. For each $\int_{CR} 1 dp_1 dp_2$ approximated, we are 99.9% confident that the approximation is within .001 units of $\int_{CR} 1 dp_1 dp_2$. This approximation is explained in more detail in the appendix. Also, we note again that the outer double sum for the computation of *EA* for each method is approximated using a Monte Carlo technique.

4. Simulation Results

We employed a Monte Carlo simulation to study the coverage and area properties of (2) – (5) for varying sample size and parameter configurations. In regards to coverage properties, we present estimated coverages for (2) – (5) across a fine grid of the unit square for the following sample size configurations: $\{(n_1, n_2) : (5,5), (10,10), (10,30), (30,30)\}$. In each sample size - parameter configuration, we randomly generated 10,000 bivariate random variables from (1). The simulation was performed in SAS IML 9.2. The nominal confidence level was 95%. In the event that a generated y_1 or y_2 was at the edges (0 or n_i), we shifted y_1 or y_2 by 1 away from the edge. Also, if a confidence region was partially outside of the parameter space, we truncated that region to only contain values in the unit square. Figures 1 – 4 summarize the coverage results.

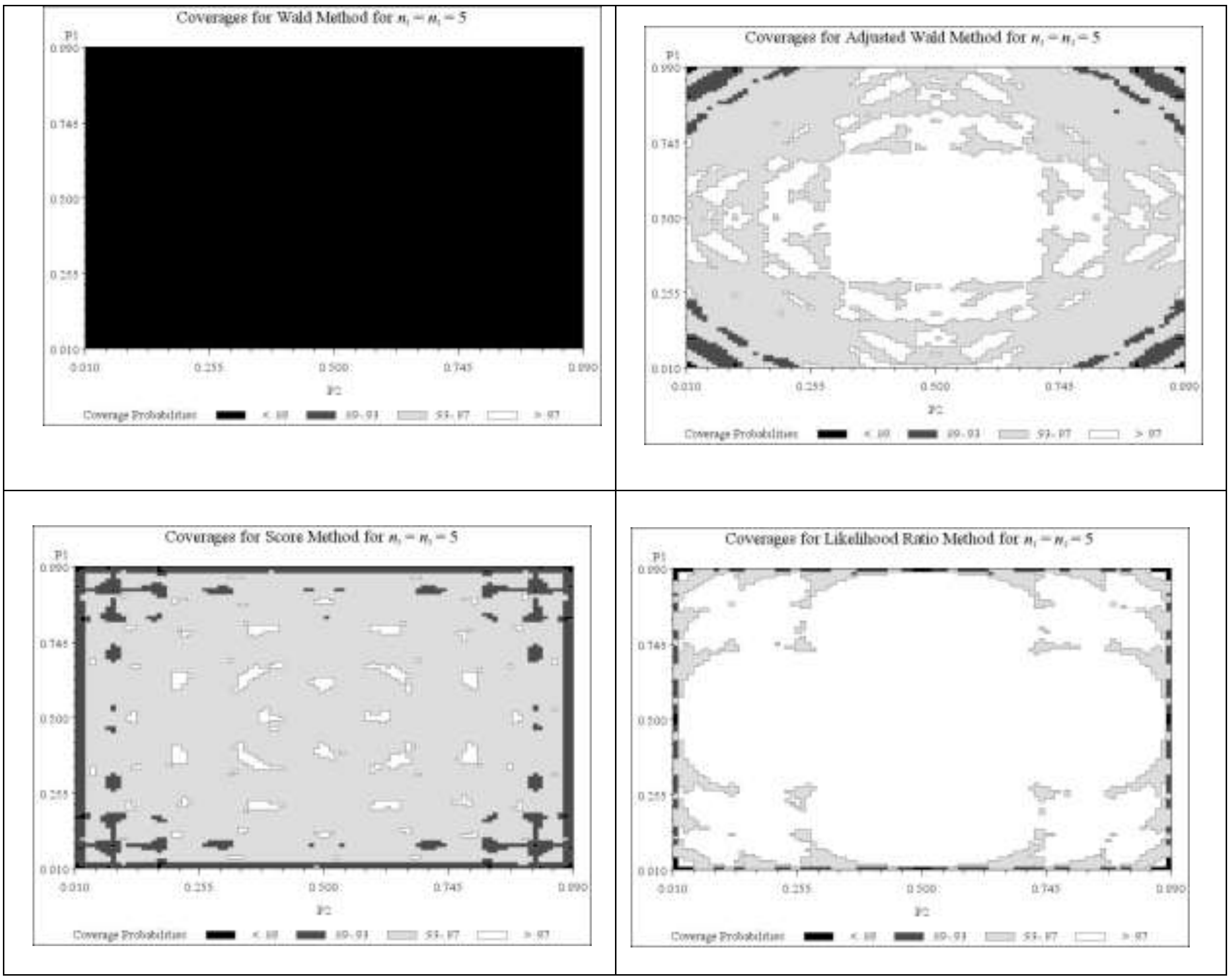


Figure 1. Estimated coverage probabilities for the Wald (W), Adjusted Wald (W*), Score (S), and Likelihood Ratio (LR) confidence regions for $n_1 = n_2 = 5$. Maximum standard error ≤ 0.005

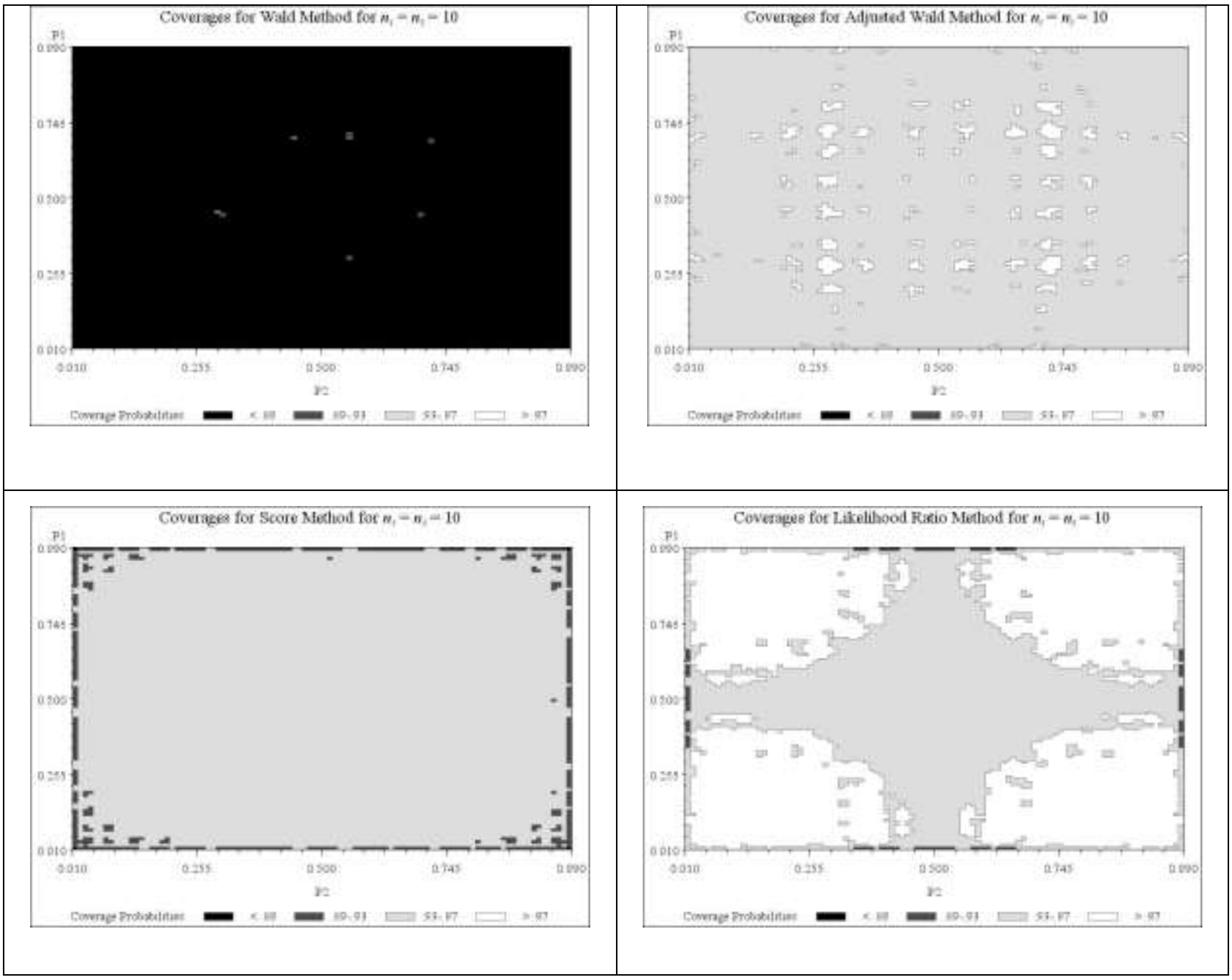
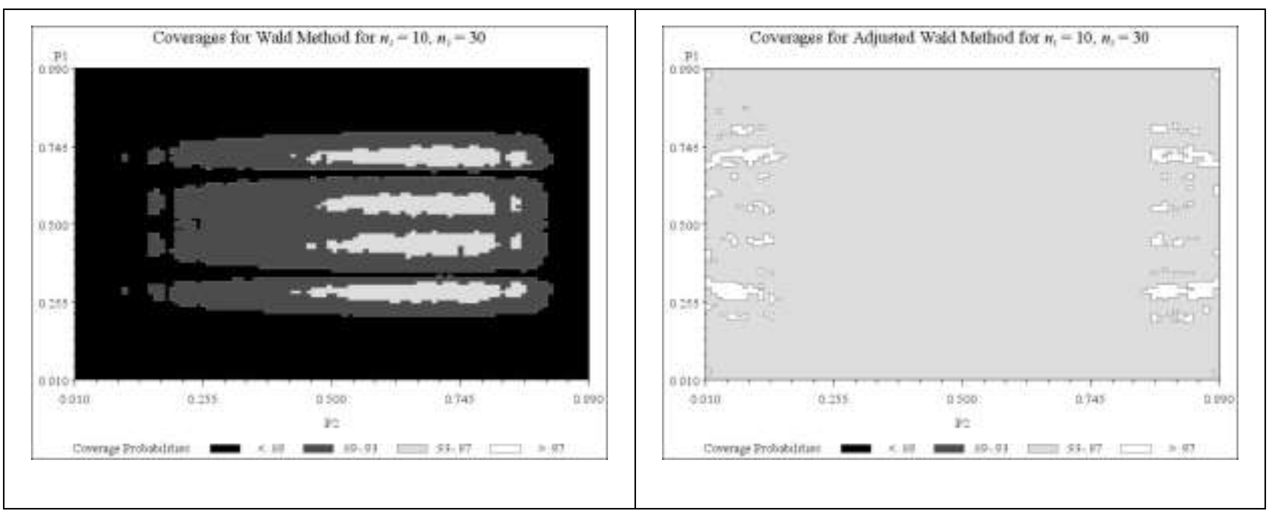


Figure 2. Estimated coverage probabilities for the Wald (W), Adjusted Wald (W*), Score (S), and Likelihood Ratio (LR) confidence regions for $n_1 = n_2 = 10$. Maximum standard error ≤ 0.005



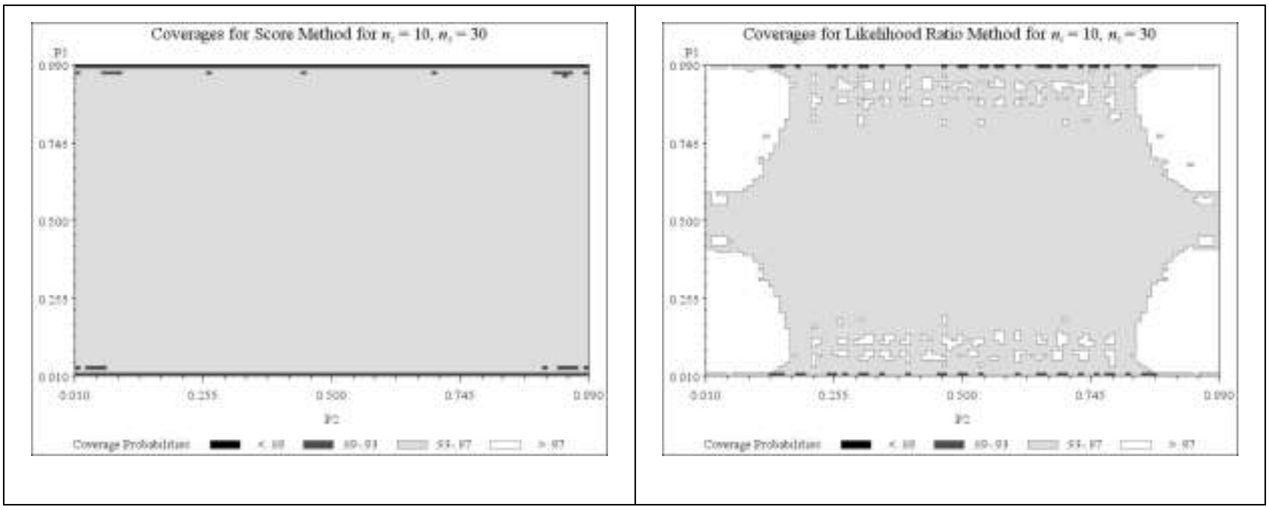


Figure 3. Estimated coverage probabilities for the Wald (W), Adjusted Wald (W*), Score (S), and Likelihood Ratio (LR) confidence regions for $n_1 = 10$ and $n_2 = 30$. Maximum standard error ≤ 0.005

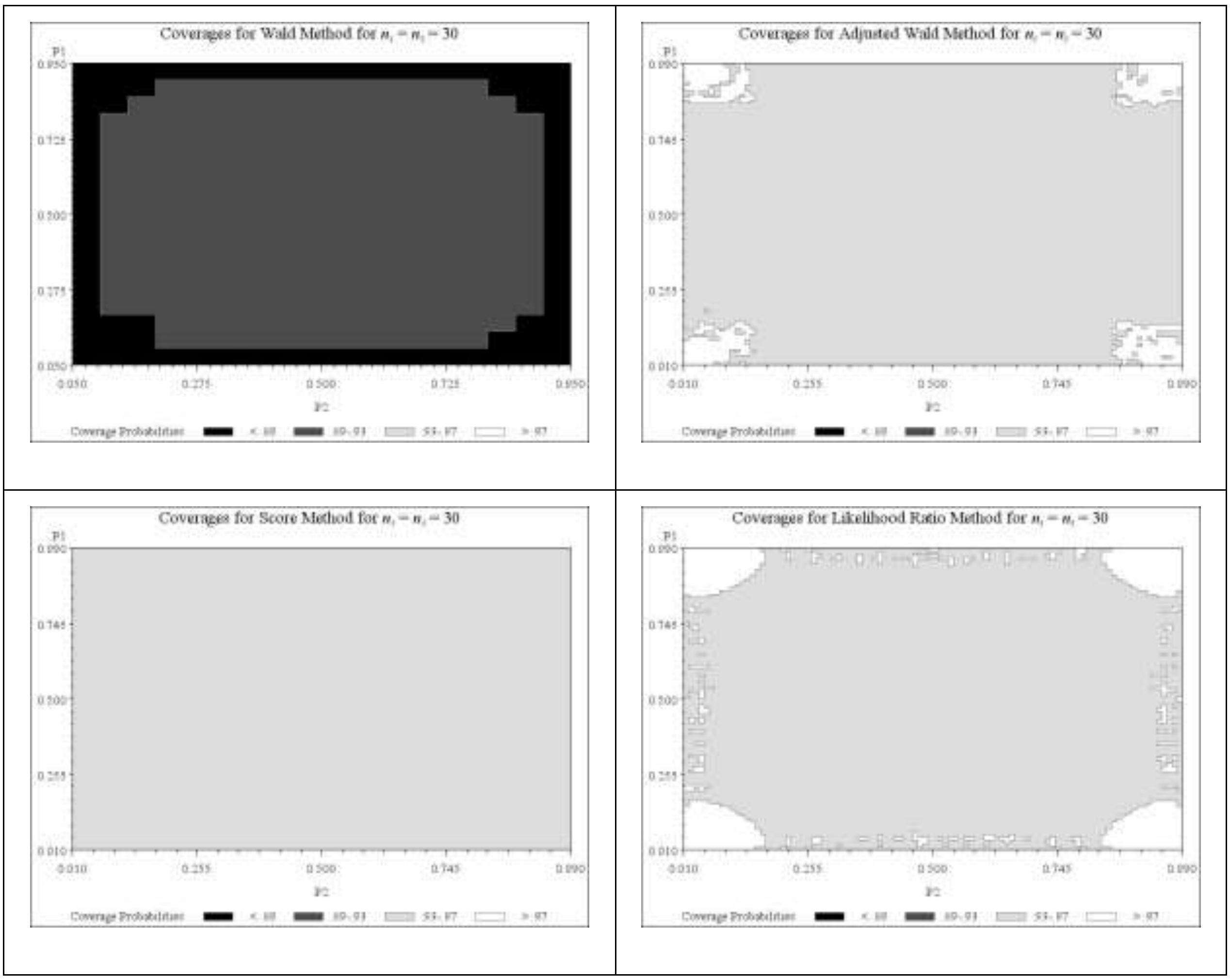


Figure 4. Estimated coverage probabilities for the Wald (W), Adjusted Wald (W*), Score (S), and Likelihood Ratio (LR) confidence regions for $n_1 = n_2 = 30$. Maximum standard error ≤ 0.005

Figures 1–4 demonstrate very poor coverage properties for the Wald ellipse. This result is consistent with other simulation studies, Agresti and Coull (1998) and Agresti and Caffo (2000), on a single proportion or difference of two proportions. Only when both sample sizes are 30 do we see the estimated coverage between 89% and 93%. Interestingly, the Wald ellipse had pockets of good coverage in the unbalanced sample-size situation where $n_1 = 10$ and $n_2 = 30$. On the other hand, the adjusted Wald procedure demonstrates good coverage properties, even when both sample sizes are as low as 5 with the exception near the corners of the unit square. Figures 1 – 4 also demonstrate that the adjusted Wald and score ellipses perform quite similarly. This result is not too surprising because, as shown in Agresti and Coull (1998), the 95% adjusted Wald interval for a single proportion is an approximation to the score interval of the same confidence level. Lastly, we point out that the likelihood-ratio ellipse has good, but often conservative coverage properties.

Table 1. Average Areas of the Wald (W), Adjusted Wald (W*), Score (S), and Likelihood Ratio (L) Methods

	$p_1 = .05$				$p_1 = .25$				$p_1 = .50$				$p_1 = .75$				$p_1 = .95$			
	W	W*	L	S	W	W*	L	S	W	W*	L	S	W	W*	L	S	W	W*	L	S
$p_2 = .05$	.367	.440	.449	.402	.421	.452	.472	.415	.467	.462	.491	.426	.424	.453	.473	.416	.367	.440	.449	.402
$p_2 = .25$	.422	.452	.472	.416	.486	.465	.497	.429	.537	.475	.517	.440	.484	.465	.496	.429	.422	.452	.472	.415
$p_2 = .50$	.468	.462	.492	.426	.537	.475	.517	.440	.598	.486	.539	.452	.536	.475	.516	.440	.467	.462	.491	.426
$p_2 = .75$	.423	.453	.473	.416	.485	.465	.496	.429	.540	.476	.518	.441	.486	.465	.497	.429	.423	.452	.473	.416
$p_2 = .95$	.367	.440	.449	.402	.421	.452	.472	.415	.468	.463	.492	.427	.422	.452	.472	.415	.367	.440	.449	.402

Note: $n_1 = 5$ and $n_2 = 5$; Maximum standard error ≤ 0.001484.

Table 2. Average Areas of the Wald (W), Adjusted Wald (W*), Score (S), and Likelihood Ratio (L) Methods

	$p_1 = .05$				$p_1 = .25$				$p_1 = .50$				$p_1 = .75$				$p_1 = .95$			
	W	W*	L	S	W	W*	L	S	W	W*	L	S	W	W*	L	S	W	W*	L	S
$p_2 = .05$	.107	.208	.171	.185	.170	.240	.215	.217	.210	.258	.242	.237	.168	.239	.214	.217	.108	.209	.171	.185
$p_2 = .25$	.169	.240	.215	.217	.264	.275	.270	.253	.328	.296	.305	.277	.265	.276	.271	.254	.169	.240	.215	.217
$p_2 = .50$	.209	.258	.242	.237	.325	.295	.303	.276	.404	.318	.343	.301	.328	.296	.305	.277	.210	.258	.242	.237
$p_2 = .75$	.169	.240	.215	.217	.267	.276	.271	.254	.327	.296	.304	.276	.265	.275	.270	.253	.169	.240	.215	.217
$p_2 = .95$	.107	.208	.170	.185	.169	.240	.215	.217	.210	.258	.243	.237	.169	.240	.215	.217	.107	.208	.171	.185

Note: $n_1 = 10$ and $n_2 = 10$; Maximum standard error ≤ 0.001004.

Table 3. Average Areas of the Wald (W), Adjusted Wald (W*), Score (S), and Likelihood Ratio (L) Methods

	$p_1 = .05$				$p_1 = .25$				$p_1 = .50$				$p_1 = .75$				$p_1 = .95$			
	W	W*	L	S	W	W*	L	S	W	W*	L	S	W	W*	L	S	W	W*	L	S
$p_2 = .05$	.048	.103	.076	.088	.076	.118	.095	.103	.094	.127	.107	.113	.076	.118	.095	.103	.048	.103	.076	.088
$p_2 = .25$	.111	.165	.135	.138	.172	.189	.169	.162	.212	.203	.190	.176	.172	.189	.169	.161	.110	.165	.134	.138
$p_2 = .50$	.129	.200	.155	.157	.200	.230	.195	.183	.247	.247	.220	.200	.200	.229	.195	.183	.129	.200	.155	.157
$p_2 = .75$	.111	.196	.134	.138	.172	.225	.169	.161	.212	.242	.191	.176	.172	.225	.169	.162	.110	.196	.134	.138
$p_2 = .95$	.048	.127	.076	.088	.076	.148	.096	.104	.094	.159	.107	.113	.076	.147	.095	.103	.048	.128	.076	.088

Note: $n_1 = 10$ and $n_2 = 30$; Maximum standard error ≤ 0.000524.

Table 4. Average Areas of the Wald (W), Adjusted Wald (W*), Score (S), and Likelihood Ratio (L) Methods

	$p_1 = .05$				$p_1 = .25$				$p_1 = .50$				$p_1 = .75$				$p_1 = .95$			
	W	W*	L	S	W	W*	L	S	W	W*	L	S	W	W*	L	S	W	W*	L	S
$p_2 = .05$	.021	.048	.033	.041	.049	.072	.059	.065	.057	.080	.069	.074	.049	.072	.059	.065	.022	.048	.033	.041
$p_2 = .25$	.049	.072	.059	.065	.112	.108	.106	.103	.129	.121	.122	.117	.111	.108	.106	.103	.049	.072	.059	.066
$p_2 = .50$	.057	.080	.069	.074	.129	.121	.122	.117	.152	.135	.141	.132	.123	.121	.122	.117	.058	.081	.069	.075
$p_2 = .75$	.049	.072	.059	.065	.111	.108	.106	.103	.129	.121	.122	.117	.111	.108	.106	.103	.049	.072	.059	.065
$p_2 = .95$	.022	.048	.033	.041	.049	.072	.059	.066	.057	.080	.069	.074	.049	.072	.059	.066	.022	.048	.033	.041

Note $n_1 = 30$ and $n_2 = 30$; Maximum standard error ≤ 0.000202.

Tables 1 – 4 give estimated average areas for (2) – (5). First, as expected, the average areas tend to decrease when the sample sizes increase. Also, from the tables we see that the Wald and score regions typically have the smallest average area for a given configuration. The Wald ellipse appears to be the smallest around the edges and corners of the unit square. This feature, though, is likely due to the fact that the Wald ellipse could have been truncated for any portion outside of the unit square, resulting in a smaller area. On the other hand, closer to the interior of the unit square, we see that the score ellipse most often has the smallest average area. We also note that the adjusted Wald ellipse displayed high average area values. This result is not too surprising given that the adjustment forced a shrinking of the p_1 and p_2 estimates towards 0.5, resulting in larger standard errors.

In summary, although the Wald ellipse often had small areas, it is not recommended because of the under-coverage problems exhibited in Figures 1 – 4. Rather, the score ellipse is preferable due to its small area values while maintaining appropriate coverage for most configurations we considered. In the case where one would want to draw a confidence ellipse by hand (e.g. introductory or intermediate statistics class), the Wald and adjusted Wald ellipses are good candidates because their construction depends only on the point estimates of p_1 and p_2, as well as the standard errors, as stated in Section 2.

5. Two Examples

Reiczigel, Abonyi, and Singer (2008) jointly estimate the sensitivity (p_1) and specificity (p_2) of an ultrasonographic examination for the bowel obstruction in dogs from data in Manczur et al. (1998). In the study, 11 of 13 dogs with mechanical ileus were correctly diagnosed while the absence of an obstruction was correctly diagnosed in 29 of 31 dogs. These diagnoses resulted in sensitivity and specificity estimates of $\hat{p}_1 = 0.85$ and $\hat{p}_2 = 0.94$. Also, as shown in Figure 5, we get the following (nominal) 95% confidence regions, which we created using the "Implicit Plot" function in Mathematica.

In addition to regions (2) – (5), a rectangular confidence set (RS) was plotted using two level $\sqrt{.95}$ % adjusted Wald confidence intervals on p_1 and p_2. Regions (2) and (3) contained values outside of the unit square, and therefore the effective regions for (2) and (3) are truncated. The (truncated, if necessary) areas for regions (2) – (5) are 0.0607, 0.1002, 0.0876, and 0.0803, respectively. The area for (2) is smaller than the areas for (3) – (5) because it was truncated to the first quadrant, not because it is inherently more precise. The area for the rectangular confidence set is 0.1082. Regions (2) and (3) and the rectangular set are symmetric, but the asymmetry of (4) and (5) is clear. Also, we note that the line $p_1 = p_2$ falls in all of the regions. Given the similar sample sizes for the simulation study resulting in Figure 3, we see that (2) is a poor choice due to its severe undercoverage, but that (3) – (5) are viable because of coverages close to the nominal level of 95%. With that in mind, the smallest confidence region here, (5), is the best choice. It has the smallest area, while maintaining appropriate coverage. It is worth noting that this likelihood-ratio confidence region is approximately 25% smaller than the rectangular set. Also, the exact or "minimum volume" confidence region from Reiczigel et al. (2008) has a slightly larger area of 0.0841 (using a grid with fin=200).

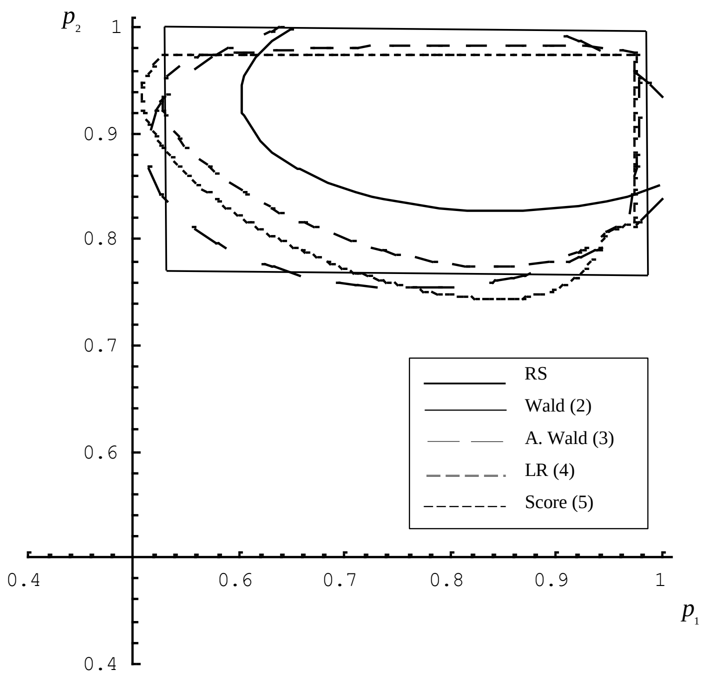


Figure 5. 95 % Confidence regions for the sensitivity and specificity of the ultrason exam

As another illustration, consider Fagerland, Lydersen, and Laake (2015), where the authors compare two survival rates, p_1 and p_2, of children who undergo a "standard" dose or "high" dose of epinephrine after an initial standard dose of epinephrine seems to be ineffective for children who remain in cardiac arrest. Perondi et al. (2004) conducted the original experiment, from which was observed 7 of 34 patients who survived at 24 hours after a standard dose, and also observed 1 of 34 who survived at 24 hours after a high dose. Of main interest for these two articles was the direct comparison of p_1 with p_2, whose point estimates are 0.21 and 0.03. A *p*-value of 0.054 using a Fisher's exact test was reported, suggesting that the high dose treatment may be worse than the standard dose treatment. Figure 6 illustrates the utility of using the bivariate confidence regions to jointly estimate p_1 and p_2. The regions allow us to see plausible values of p_1 and p_2, while at the same time testing their equality with a graphical check of the line. As before, we created the following (nominal) 95% confidence regions with the “Implicit Plot” function in Mathematica.

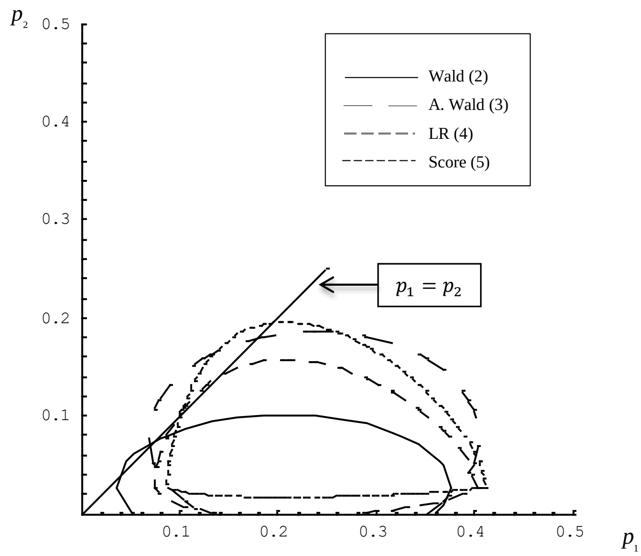


Figure 6. 95% Confidence regions for survival rates using the standard dose and high dose, respectively, for epinephrine study; along with line to test H_0: $p_1 = p_2$

6. Discussion

Other than the afore-mentioned Reiczigel et al. (2008) article, the authors found no published method for joint estimation of two binomial proportions. Bivariate confidence regions are useful because they describe where the individual proportions might lie, how they interact, and how they compare to one another, as was illustrated in two different examples. The first example’s main focus was joint *estimation* of the sensitivity and specificity of a diagnostic test. The second example’s main objective was the *comparison* of two survival rates but with the added inferential feature of plausible values for the individual survival rates. The findings of this paper should prove useful for applied statisticians who desire to jointly estimate binomial proportions from independent samples. We also contend that the findings here are useful for elementary and intermediate statistics instructors when they cover statistical inference on two population proportions.

Some future work in this area is to consider a “Clopper-Pearson” based confidence region for the two proportions, as well as confidence regions for two binomial proportions when the two samples are dependent. One could also extend the methods here to three proportions, and give a three-dimensional confidence region. Beyond three dimensions, though, the utility of graphically seeing the confidence region is greatly diminished.

Acknowledgements

This work is supported, in part, by a research grant (107552-26413-150) and publication support from the Office of Research and Sponsored Programs at Stephen F. Austin State University, as well as the Department of Mathematics and Statistics at SFASU. The authors also thank Baylen Johnson and Kacey McCauley for their assistance in the simulation study, as well as Mrs. Joy Young for her editorial work on the manuscript. No potential competing interest is reported by the authors.

References

Agresti, A., & Coull, B. (1998). Approximate is Better Than ‘Exact’ for Interval Estimation of Binomial Proportions. *The American Statistician*, *52*, 119-126. https://doi.org/10.1080/00031305.1998.10480550

Agresti, A., & Caffo, B. (2000). Simple and Effective Confidence Intervals for Proportions and Differences of Proportions Result from Adding Two Successes and Two Failures. *The American Statistician*, *54*, 280-288. https://doi.org/10.1080/00031305.2000.10474560

Elliot, P., & Riggs, K. (2015). Confidence Regions for Two Proportions from Independent Negative Binomial Distributions. *Journal of Applied Statistics*, *42*(1), 27-36. https://doi.org/10.1080/02664763.2014.929642

Fagerland, M., Lydersen, S., & Laake, P. (2015). Recommended Confidence Intervals for Two Independent Binomial Proportions. *Statistical Methods in Medical Research*, *24*(2), 224-254. https://doi.org/10.1177/0962280211415469

Manczur, F., Voros, K., Vrabley, T., Wladar, S., Nemeth, T., & Fenyves, B. (1998). Sonographic Diagnosis of Intestinal Obstruction in the Dog. *Acta Veterinaria Hungarica*, *46*(1), 35-45.

Newcombe, R. (1998a). Two-sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods," *Statistics in Medicine*, *17*, 857-872. https://doi.org/10.1002/(SICI)1097-0258(19980430)17:8<857::AID-SIM777>3.0.CO;2-E

Newcombe, R. (1998b). Interval Estimation for the Difference Between Independent Proportions: Comparison of Eleven Methods. *Statistics in Medicine*, *17*, 873-890. https://doi.org/10.1002/(SICI)1097-0258(19980430)17:8<857::AID-SIM777>3.0.CO;2-E

Perondi, M., Reis, A., Paiva, E., Nadkarni, V., & Berg, R. (2004). A Comparison of High-dose and Standard-dose Epinephrine in Children with Cardiac Arrest," *The New England Journal of Medicine*, *350*, 1722-1730. https://doi.org/10.1056/NEJMoa032440

Reiczigel, J., Abonyi-Toth, Z., & Singer, J. (2008). An Exact Confidence Set for Two Binomial Proportions and Exact Unconditional Confidence Intervals for the Difference and Ratio of Proportions. *Computational Statistics and Data Analysis*, *52*, 5046-5053. https://doi.org/10.1016/j.csda.2008.04.032

Schilling, M., & Doi, J. (2014). A Coverage Probability Approach to Finding an Optimal Binomial Confidence Procedure. *The American Statistician*, *68*, 133–145. https://doi.org/10.1080/00031305.2014.899274

Viana, M. (1991). Bayesian Joint Estimation of Binomial Proportions. *Journal of Educational Statistics*, *16*(4), 331-343. https://doi.org/10.3102/10769986016004331

Appendix

One method of approximating $\int_{CR} 1dp_1dp_2$ involves randomly generating two independent uniform variates on the interval [0, 1] and then checking to see if the pair lies in the confidence region, i.e. satisfies the pertinent inequality in (2), (3), (4), or (5). The number of pairs that land in the region, *w*, divided by the total number of pairs generated, *m*, acts as an approximation of $\int_{CR} 1dp_1dp_2$. The challenge here is finding a suitable value for *m* so that w/m is a good approximation of $\int_{CR} 1dp_1dp_2$. Our desire was for w/m to be within 0.001 of $\int_{CR} 1dp_1dp_2$ with 99.9% confidence. To achieve this goal, we choose *m* such that $m = \frac{(3.29)^2 A_{CR}(1-A_{CR})}{(0.001)^2}$ where A_{CR} is the area of the confidence region and 3.29 is the 99.95th percentile of the standard normal distribution. Except in the case of (2), A_{CR} is obviously unknown, however, we can choose a conservative value of A_{CR} that ensures our approximations are still within 0.001 of $\int_{CR} 1dp_1dp_2$ with 99.9% confidence. The most conservative choice of A_{CR} is 0.5, which yields $m = 2{,}706{,}025$, but this is computationally inefficient and unnecessary. Because (2) is an ellipse, we know that $A_{CR} = 4\pi\chi^2_{1-\alpha}(2)\hat{\sigma}_{\hat{p}_1}\hat{\sigma}_{\hat{p}_2}$ for the Wald-based approach (2). Also, computational experience suggests that the difference in the four regions' areas is no more than 0.1. Therefore, we recommend using the value $4\pi\chi^2_{1-\alpha}(2)\hat{\sigma}_{\hat{p}_1}\hat{\sigma}_{\hat{p}_2} + 0.1$ as a suitable plug-in for A_{CR} because $4\pi\chi^2_{1-\alpha}(2)\hat{\sigma}_{\hat{p}_1}\hat{\sigma}_{\hat{p}_2}$ is considerably less than 0.5, and the addition of 0.1 is a conservative shrinking towards 0.5.