Journal of the Korean Statistical Society (2007), **36**: *2, pp 275–284*

AN IMPROVED CONFIDENCE INTERVAL FOR THE POPULATION PROPORTION IN A DOUBLE SAMPLING SCHEME SUBJECT TO FALSE-POSITIVE MISCLASSIFICATION†

SEUNG-CHUN LEE[1]

ABSTRACT

Confidence intervals for the population proportion in a double sampling scheme subject to false-positive misclassification are considered. The confidence intervals are obtained by applying Agresti and Coull's approach, so-called "adding two-failures and two successes". They are compared in terms of coverage probabilities and expected widths with the Wald interval and the confidence interval given by Boese *et al.* (2006). The latter one is a test-based confidence interval and is known to have good properties. It is shown that the Agresti and Coull's approach provides a relatively simple but effective confidence interval.

AMS 2000 subject classifications. Primary 62F25; Secondary 62D05.
Keywords. Agresti-Coull confidence interval, confidence interval, double sampling, false-positive misclassification.

1. INTRODUCTION

A double sampling scheme occurs when the cost of precise test is expensive. To reduce the cost, a large sample is classified by an inexpensive but fallible device, and a subsample is classified by a supplementary inerrant device. A large amount of literature concerned about the inference on the population proportion in the double sampling (for example, Tenenbein, 1970, 1971, 1972; Geng and Asano, 1989; York *et al.*, 1995; Moors *et al.*, 2000; Barnett *et al.*, 2001; Raats and Moors, 2003; Boese *et al.*, 2006). For instance, York *et al.* (1995) estimated

Received August 2006; accepted January 2007.
†This work was supported by Hanshin University Research Grant in 2007.
[1]Department of Statistics, Hanshin University, Osan, Kyunggi-Do 447-791, Korea (e-mail: seung@hs.ac.kr)

the proportion of infants born with Down's syndrome in Norway with a double sampling data. Their data was collected by two stages. At the first stage, for every birth during a certain time period, the midwife or obstetrician classified the child with Down's syndrome based on a visual inspection. Because the classification by a visual inspection could not be expected to be accurate, the data might be exposed to measurement error. Since Bross (1954) warned that usual estimators could be extremely biased when data is subject to misclassification, it might dangerous to use the data alone. An additional data would be required and supplied by the second stage. A small subsample of births was classified by accurate but expensive cytogenetic test. A double sampling data was formed by these two stages.

The subsample might contain two types of error done by the midwife or obstetrician. He or she might classify erroneously a normal child into Down's syndrome (false-positive) and vise versa (false-negative). Tenenbein (1970) presented the maximum likelihood (ML) estimator for the population proportion as well as for false-positive and false-negative error rates (Tenenbein, 1971, 1972). The same ML estimator for the population proportion was derived by Barnett *et al.* (2001), and was applied to a specific data analysis.

Some authors have considered the model with only one type of misclassification. For example, Lie *et al.* (1994) consider the case that the false-negative counts were corrected using multiple fallible classifiers and gave the ML estimators. The same model was considered by York *et al.* (1995). They estimated the proportion of Down syndrome in Norway from the Bayesian prospective. Moors *et al.* (2000) discussed one-sided interval estimation in the case that only false-negative misclassification occurs. This result was generalized by Raats and Moors (2003) to the case with both types of misclassification. They obtained a Bayesian posterior distribution using a conjugated prior. The posterior distribution was utilized to construct a one-sided confidence interval.

For the interval estimation problem, Boese *et al.* (2006) gave asymptotic confidence intervals in the false-positive misclassification model. Although the intervals are based on standard frequentist methods, certain combinations of likelihood, Fisher-information types and likelihood-based statistics, the performance of the intervals is remarkable in that the coverage probabilities are much close to nominal level even with a small sample. Another interesting feature of the intervals is that, they claimed, the intervals were the first two-sided frequentist confidence intervals for the population proportion. However, these are test-based confidence intervals and it is hard to compute the confidence limits. The limits can be computed only numerically. One might prefer confidence intervals with closed form.

Recently the Wald intervals for proportion or difference of two proportions adjusted by adding two failures and two successes were shown to have remarkably good properties compared with the original Wald intervals (Agresti and Coull, 1998; Agresti and Caffo, 2000; Price and Bonett, 2004; Agresti and Min, 2005). More importantly, they are simple but comparable with other widely used confidence intervals. It is recognized by many authors (Brown *et al.*, 2001; Agresti and Coull, 1998; Lee, 2006a, 2006b) that the simplicity is an important factor for practical usage.

In this note, confidence intervals based on Agresti-Coull's approach are considered and the coverage probabilities and the expected widths of the intervals are compared. Since the confidence interval based on full-likelihood score statistic, which was given by Boese *et al.* (2006) and denoted by $\mathrm{CI}\{S_n[I_{pp|\phi}]\}$, is known to have good properties, it is included in this comparison as a referential interval. One might refer to Boese *et al.* (2006) for the performance of $\mathrm{CI}\{S_n[I_{pp|\phi}]\}$. However, they estimated the coverage probability and the expected width by a Monte Carlo simulation. It seems that the estimate is well done enough to capture the property of intervals, but it might be desirable to compute exact coverage probabilities and expected widths for the precise comparison. Thus, we recalculate the properties of $\mathrm{CI}\{S_n[I_{pp|\phi}]\}$.

2. CONFIDENCE INTERVALS IN THE FALSE-POSITIVE MODEL

The model considered in this paper is the same that of Boese *et al.* (2006) and in what follows we will use the notation of Section 2 of Boese *et al.* (2006).

As we noted before, a double sampling scheme consists of two stages of sampling. A sample of size N is selected at random from the population of interest and each unit in the sample is classified by a fallible device. And then a subset of size n is selected from the initial sample. Each unit in the subsample is tested by an inerrant device. Thus, a unit in subsample is tested by both the inerrant and the fallible device.

For each unit tested by the inerrant device, let $T_i = 1$, if the i^{th} unit is recorded to be positive (or a success) with the inerrant device, and $T_i = 0$, if otherwise. Likely, for each unit tested by the fallible device, define $F_i = 1$, if the i^{th} unit is classified into positive with the fallible device, and $F_i = 0$, if otherwise. Because the sample is assumed to be selected at random from an infinite population, the proportion of success p can be written as

$$p = \Pr[\, T_i = 1 \,]$$

and the false-positive error rate is defined to be

$$\phi = \Pr[\, F_i = 1 \mid T_i = 0 \,].$$

The false-negative error rate $\Pr[F_i = 0 | T_i = 1]$ is assumed to zero in this model. Thus each unit in the subsample belongs to one of three mutually disjoint categories $\{(t, f) | \ (0,0), (0,1), (1,1)\}$ with probabilities $(1-p)(1-\phi), (1-p)\phi$ and p, respectively. Let n_{tf} be the observed counts in (t, f).

$N - n$ units are tested by only fallible device. Among these units, let x be the number of counts that test positive and $y = N - n - x$. Note that $N = n + x + y, n = n_{00} + n_{01} + n_{11}$ and $\Pr[F_i = 1] = p + (1-p)\phi$.

Since each unit is tested independently, the joint likelihood of p and ϕ is

$$L(p, \phi) = A[(1-p)(1-\phi)]^{n_{00}}[(1-p)\phi]^{n_{01}} p^{n_{11}} \pi^x (1-\pi)^y, \tag{2.1}$$

where $A = n!/(n_{00}! n_{01}! n_{11}!)\binom{N-n}{x}$ and $\pi = p + (1-p)\phi$. For model (2.1), Tenenbein (1970) gave the maximum likelihood estimate of p,

$$\hat{p} = \frac{n_{11}}{n_{01} + n_{11}} \frac{x + n_{01} + n_{11}}{N} \tag{2.2}$$

and the approximation of its variance,

$$\widehat{\mathrm{Var}}(\hat{p}) = \frac{\hat{p}\,\hat{q}}{n}\left[1 - \frac{\hat{p}\,\hat{q}(1-\hat{\phi})^2}{\hat{\pi}(1-\hat{\pi})}\right] + \frac{\hat{p}^2\hat{q}^2(1-\hat{\phi})^2}{N\hat{\pi}(1-\hat{\pi})}, \tag{2.3}$$

where $\hat{q} = 1 - \hat{p}, \hat{\pi} = \hat{p} + (1-\hat{p})\hat{\phi}$ and

$$\hat{\phi} = \frac{n_{01}}{n_{01} + n_{11}} \frac{x + n_{01} + n_{11}}{N(1-\hat{p})}$$

(Boese *et al.*, 2006). Note that (2.3) can be further simplified as

$$\widehat{\mathrm{Var}}(\hat{p}) = \frac{\hat{p}\,\hat{q}}{n} - \left(\frac{1}{n} - \frac{1}{N}\right) \frac{n_{11}}{n_{11} + n_{01}} \hat{p}(1-\hat{\pi}). \tag{2.4}$$

Using (2.2) and (2.4), we have an asymptotic confidence interval for p, which is given by

$$\mathrm{CI}\{\hat{p}\} = \hat{p} \pm z_{\alpha/2}\sqrt{\widehat{\mathrm{Var}}(\hat{p})} \tag{2.5}$$

and is called Wald interval.

The confidence interval based on Agresti and Coull's approach is exactly the same that of (2.5) with adjusted observed counts, adding 4 artificial observations. However, in this case it is not clear how to add pseudo observations. Since $1 - \pi = (1 - p)(1 - \phi)$, one may regard (2.1) as the likelihood of data with 4 categories. Then add 1 pseudo observation to each category, which would lead to

$$\bar{n}_{00} = n_{00} + \frac{1}{2},\ \bar{n}_{01} = n_{01} + 1,\ \bar{n}_{11} = n_{11} + 1,\ \bar{x} = x + 1,\ \bar{y} = y + \frac{1}{2}. \tag{2.6}$$

The total number of observation and the size of subsample become $\bar{N} = N + 4$ and $\bar{n} = n + 5/2$, respectively. On the other hand, it also could be considered as the product of two binomial likelihoods. Add two failures and two successes for each binomial count, then

$$\tilde{n}_{00} = n_{00} + 1,\ \tilde{n}_{01} = n_{01} + 1,\ \tilde{n}_{11} = n_{11} + 2,\ \tilde{x} = x + 2,\ \tilde{y} = y + 2 \tag{2.7}$$

and we have $\tilde{N} = N + 8, \tilde{n} = n + 4$.

Applying these two sets of artificial data to (2.2) and (2.4), we have two Agresti-Coull type confidence intervals which will be denoted by CI$\{\bar{p}\}$ and CI$\{\tilde{p}\}$, respectively. In Section 3, we will examine the property of these two intervals compared with Wald interval and CI$\{S_n[I_{pp|\phi}]\}$.

3. Comparison Results and Conclusion

The effect of pseudo observations is getting smaller as n and N are increasing. Thus we will examine the property of two Agresti-Coull type intervals in small samples, say $N = 30$ to 400. The size of subsample is assumed to be 10% of the total sample.

Figure 3.1 shows the coverage probabilities and the expected widths of four 95% confidence intervals when $p = \phi = 0.1$. The values shown in the figures were computed on every 10 from $N = 30$ to 400 using R 2.1.1. Obviously Wald interval has many problems in small samples. For instance, the rate of convergence for CI$(\hat{p})$ is considerably slower than the other three. Also it should be noted that Wald interval has a considerably smaller coverage probability with similar value of width when N is relatively large, say $N = 400$. Thus we can conclude that the center of CI$(\hat{p})$ is wrong. This might justify adjusting $\hat{p}$ somehow.

Beside Wald interval, the other three intervals are seemed to converge toward the nominal confidence level as N increases. Among the three intervals,

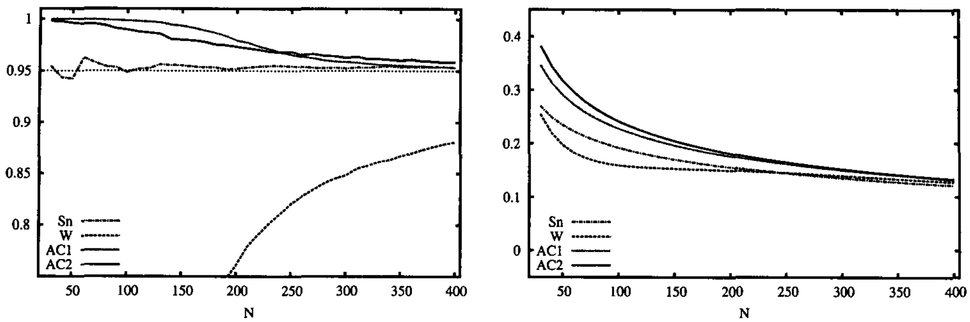

FIGURE 3.1 *Coverage probabilities (left) and expected widths (right) of 95% confidence intervals for* $p = \phi = 0.1, N = 30$ *to* 400 *and* $n = 0.1N$. *"Sn", "W", "AC1" and "AC2" represent* $CI\{S_n[I_{pp|\phi}]\}$, $CI(\hat{p})$, $CI(\bar{p})$ *and* $CI(\tilde{p})$, *respectively.*

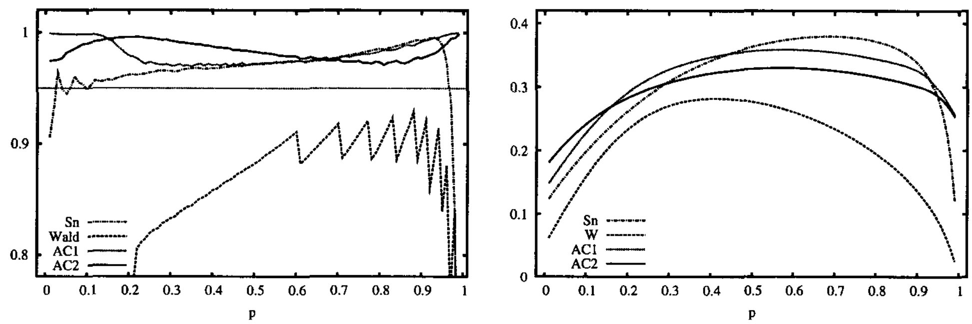

FIGURE 3.2 *Coverage probabilities (left) and expected widths (right) of 95% confidence intervals for* $p = 0.01$ *to* 0.99 *with* $\phi = 0.1, N = 100$ *and* $n = 0.1N$. *"Sn", "W", "AC1" and "AC2" represent* $CI\{S_n[I_{pp|\phi}]\}$, $CI(\hat{p})$, $CI(\bar{p})$ *and* $CI(\tilde{p})$, *respectively.*

CI$\{S_n[I_{pp|\phi}]\}$ has the closest coverage to the nominal level, which shows it has the best characteristics with respect to approximation. However, Boese *et al.* (2006) noted that CI$\{S_n[I_{pp|\phi}]\}$ has its best performance when p is small, say $p = 0.1$ to 0.25. In fact, we found that as p goes to 1, the coverage of CI$\{S_n[I_{pp|\phi}]\}$ approaches to 1, but is dropped suddenly at very near 1 when N is small. Figure 3.2 depicts this phenomenon. Apparently, we could not say that CI$\{S_n[I_{pp|\phi}]\}$ would be uniformly better than CI$(\bar{p})$ and CI$(\tilde{p})$ with respect to the approximation.

The width of interval is another important property for judgment. In Figure 3.1, CI$(\tilde{p})$ has the widest expected width, but the differences in the widths are getting smaller as N increases. The disparity in the width might be negligible when N is large, say greater than 300. Note that when $N = 300$, the size of subsample is only 30. Thus the differences are not significant from a practical

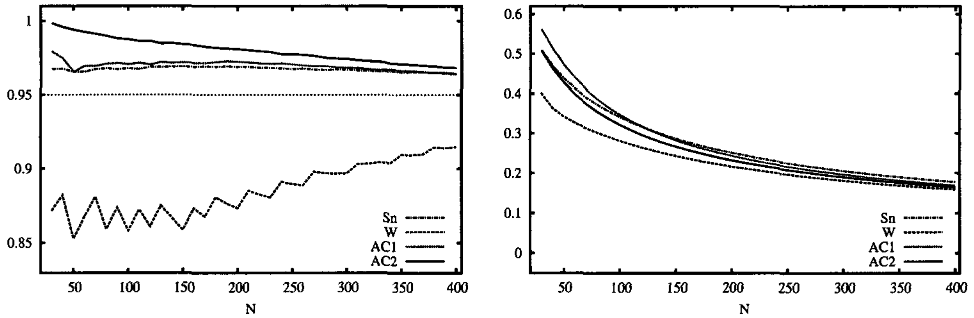

FIGURE 3.3 *Coverage probabilities (left) and expected widths (right) of 95% confidence intervals for* $p = 0.4, \phi = 0.1, N = 30$ *to* 400 *and* $n = 0.1N$. *"Sn", "W", "AC1" and "AC2" represent* $CI\{S_n[I_{pp|\phi}]\}$, $CI(\hat{p})$, $CI(\bar{p})$ *and* $CI(\tilde{p})$, *respectively.*

point of view. Nonetheless, it seems that the Agresti-Coull type intervals are unnecessarily wide. This is a common feature of intervals based on Agresti and Coull's approach (Agresti and Coull, 1998; Agresti and Caffo, 2000; Lee, 2006a, 2006b).

With the same configuration of parameter values of Figure 3.1 except $p = 0.4$, we get Figure 3.3. Again Wald interval behaves poorly. The coverage probabilities of CI($\bar{p}$) and CI$\{S_n[I_{pp|\phi}]\}$ are smaller and closer than that of CI($\tilde{p}$) to the nominal level. Note however, they have wider expected widths than CI($\tilde{p}$). It is not intuitively clear, but the result coincides with Figure 3.2. Perhaps, $\tilde{p}$ is a good choice for the center of interval. We might prefer CI($\tilde{p}$) to other intervals in this configuration, because CI($\tilde{p}$) have a large value of coverage with narrow interval width.

Based on these observations, we may conclude that both CI($\bar{p}$) and CI($\tilde{p}$) are at least practically comparable with CI$\{S_n[I_{pp|\phi}]\}$ in terms of their performances, and are better in simplicity. However, we could not conclude which one is better choice among two Agresti-Coull type intervals. At the first glance, CI($\tilde{p}$) seems to be better. However, two intervals behave very complicatedly. To reach a conclusion, more close examination is required. Thus with various combinations of parameter values, $N = 100, 200, 300, 400$, $\phi = 0.1, 0.2, 0.3$ and $n = 0.1N, 0.2N, 0.3N$, the coverage probability and the expected width of intervals for every $p = 0.01$ to 0.99 were obtained to calculate the average of coverage probabilities, the root mean square deviation (RMSD) from the nominal confidence level and the average of expected widths. The average of expected width of CI$\{S_n[I_{pp|\phi}]\}$ is not considered in this calculation. The reason for the exclusion is that it is computationally too burden.

The result is tabulated in Table 3.1. Because Wald interval behaves very

TABLE 3.1 *Average and RMSD of coverage probabilities, and average expected width*

			Mean Coverage(RMSD)				*Mean Expected Width*		
ϕ	n	N	$CI\{S_n\}$	$CI\{\hat{p}\}$	$CI\{\bar{p}\}$	$CI\{\tilde{p}\}$	$CI\{\hat{p}\}$	$CI\{\bar{p}\}$	$CI\{\tilde{p}\}$
0.1	$0.1N$	100	0.965(.044)	0.816(.186)	0.981(.033)	0.983(.034)	0.212	0.315	0.298
		200	0.971(.028)	0.846(.150)	0.978(.029)	0.977(.028)	0.166	0.214	0.208
		300	0.972(.026)	0.874(.116)	0.974(.026)	0.972(.024)	0.142	0.170	0.167
		400	0.972(.026)	0.892(.093)	0.970(.023)	0.969(.021)	0.125	0.145	0.143
	$0.2N$	100	0.968(.046)	0.872(.138)	0.970(.022)	0.963(.018)	0.195	0.234	0.229
		200	0.970(.029)	0.908(.087)	0.964(.017)	0.958(.010)	0.144	0.160	0.159
		300	0.970(.027)	0.921(.065)	0.961(.014)	0.956(.008)	0.119	0.128	0.128
		400	0.970(.026)	0.928(.048)	0.959(.012)	0.955(.006)	0.104	0.110	0.110
	$0.3N$	100	0.966(.046)	0.898(.110)	0.963(.016)	0.952(.017)	0.182	0.202	0.201
		200	0.969(.029)	0.925(.062)	0.958(.011)	0.951(.012)	0.132	0.140	0.140
		300	0.971(.027)	0.932(.044)	0.956(.009)	0.951(.007)	0.109	0.113	0.113
		400	0.971(.028)	0.937(.030)	0.955(.008)	0.950(.007)	0.094	0.097	0.097
0.2	$0.1N$	100	0.964(.049)	0.755(.235)	0.977(.029)	0.974(.029)	0.255	0.340	0.324
		200	0.971(.030)	0.837(.154)	0.970(.025)	0.968(.022)	0.204	0.239	0.233
		300	0.972(.027)	0.869(.120)	0.966(.023)	0.965(.019)	0.173	0.193	0.190
		400	0.972(.027)	0.887(.098)	0.964(.021)	0.963(.017)	0.153	0.166	0.165
	$0.2N$	100	0.965(.050)	0.859(.147)	0.967(.020)	0.960(.013)	0.223	0.254	0.249
		200	0.970(.031)	0.902(.089)	0.961(.017)	0.957(.008)	0.166	0.177	0.176
		300	0.972(.028)	0.917(.067)	0.959(.015)	0.956(.007)	0.137	0.144	0.143
		400	0.972(.027)	0.924(.050)	0.957(.013)	0.955(.006)	0.120	0.124	0.124
	$0.3N$	100	0.965(.050)	0.892(.114)	0.961(.016)	0.952(.011)	0.201	0.217	0.216
		200	0.970(.031)	0.921(.063)	0.957(.012)	0.951(.008)	0.146	0.152	0.152
		300	0.971(.028)	0.930(.046)	0.955(.010)	0.952(.005)	0.120	0.124	0.124
		400	0.972(.028)	0.935(.032)	0.954(.009)	0.951(.005)	0.105	0.107	0.107
0.3	$0.1N$	100	0.962(.055)	0.746(.244)	0.972(.025)	0.967(.025)	0.290	0.360	0.346
		200	0.970(.033)	0.836(.160)	0.965(.023)	0.963(.019)	0.230	0.258	0.253
		300	0.972(.028)	0.868(.126)	0.962(.021)	0.961(.017)	0.195	0.211	0.208
		400	0.972(.027)	0.886(.104)	0.960(.019)	0.959(.016)	0.172	0.182	0.181
	$0.2N$	100	0.963(.055)	0.855(.152)	0.964(.020)	0.958(.011)	0.245	0.270	0.266
		200	0.969(.033)	0.899(.092)	0.959(.017)	0.957(.009)	0.181	0.191	0.190
		300	0.971(.028)	0.914(.070)	0.957(.015)	0.955(.008)	0.150	0.155	0.155
		400	0.971(.027)	0.923(.054)	0.956(.013)	0.955(.007)	0.131	0.134	0.134
	$0.3N$	100	0.962(.055)	0.888(.117)	0.960(.016)	0.953(.007)	0.216	0.229	0.228
		200	0.969(.033)	0.919(.065)	0.956(.013)	0.952(.006)	0.157	0.162	0.162
		300	0.970(.028)	0.928(.048)	0.955(.011)	0.952(.004)	0.129	0.132	0.132
		400	0.971(.028)	0.934(.034)	0.954(.009)	0.952(.004)	0.112	0.114	0.114

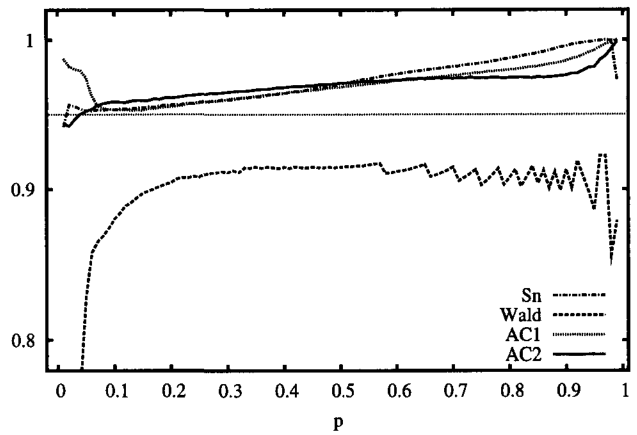

FIGURE 3.4 *Coverage probabilities for* $p = 0.01$ *to* 0.99 *with* $\phi = 0.1, N = 400$ *and* $n = 0.1N$. *"Sn", "W", "AC1" and "AC2" represent* $CI\{S_n[I_{pp|\phi}]\}$, $CI(\hat{p})$, $CI(\bar{p})$ *and* $CI(\tilde{p})$, *respectively.*

poorly compared with others, we may safely exclude it in what follow. Note that among the three, CI($\tilde{p}$) has the closest average coverage probability to nominal level 95% in most cases, and always has smaller average expected width than CI($\bar{p}$). Apparently CI($\tilde{p}$) outperforms and hence is preferable to CI($\bar{p}$) in average sense.

Note also, it is intuitively clear that the average coverage probability should be approaching to the nominal level as N or n increases. Two Agresti-Coull type intervals show this pattern. However, CI$\{S_n[I_{pp|\phi}]\}$ disagrees with the pattern. Although the RMSD of CI$\{S_n[I_{pp|\phi}]\}$ decreases as N gets increasing, the best approximation occurs when N is small. The increment of N or n does not necessarily improve the approximation in average sense. One might be wondering about this fact because it is based on a novel large sample theory. We believe that it might be due to the fact that as you can see in Figure 3.2, with small N, the coverage probability is suddenly dropped toward 0 as p approaches to 1. This would make the average coverage small to fit the nominal level with large RMSD. Figure 3.4 demonstrates that the size of drop is reduced by enlarging N, which in turn enlarges the average coverage probability but reduce RMSD. As you can see, CI$\{S_n[I_{pp|\phi}]\}$ has normally larger coverage than nominal level. This is true for CI($\bar{p}$) and CI($\tilde{p}$) too, but CI($\tilde{p}$) gives better approximation in average sense. This might justify the usability of CI($\tilde{p}$). However, it is desirable to reduce the width of CI($\tilde{p}$) somehow. We believe that a noninformative Bayes approach might do work well as in Lee (2006a, 2006b).

REFERENCES

AGRESTI, A. AND COULL, B. A. (1998). "Approximate is better than "exact" for interval estimation of binomial proportions", *The American Statistician,* **52**, 119–126.

AGRESTI, A. AND 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.

AGRESTI, A. AND MIN, Y. (2005). "Simple improved confidence intervals for comparing matched proportions", *Statistics in Medicine,* **24**, 729-740.

BARNETT, V., HAWORTH, J. AND SMITH, T. M. F. (2001). "A two-phase sampling scheme with applications to auditing or *sed quis custodiet ipsos custodes?*", *Journal of Royal Statistical Society,* Ser. A, **164**, 407–422.

BOESE, D. H., YOUNG, D. M. AND STAMEY, J. D. (2006). "Confidence intervals for a binomial parameter based on binary data subject to false-positive misclassification", *Computational Statistics & Data Analysis,* **50**, 3369–3385.

BROSS, I. (1954). "Misclassification in 2 × 2 tables", *Biometrics,* **10**, 478–486.

BROWN, L. D., CAI, T. T. AND DASGUPTA, A. (2001). "Interval estimation for a binomial proportion", *Statistical Science,* **16**, 101–133.

GENG, Z. AND ASANO, C. (1989). "Bayesian estimation methods for categorical data with misclassifications", *Communications in Statistics-Theory and Methods,* **18**, 2935–2954.

LEE, S.-C. (2006a). "The weighted Polya posterior confidence interval for the difference between two independent proportions", *The Korean Journal of Applied Statistics,* **19**, 171–181.

LEE, S.-C. (2006b). "Interval estimation of binomial proportions based on weighted Polya posterior", *Computational Statistics & Data Analysis.* **51**, 1012–1021.

LIE, R. T., HEUCH, I. AND IRGENS, L. M. (1994). "Maximum likelihood estimation of the proportion of congenital malformations using double registration systems", *Biometrics,* **50**, 433-444.

MOORS, J. J. A., VAN DER GENUGTEN, B. B. AND STRIJBOSCH, L. W. G. (2000). "Repeated audit controls", *Statistica Neerlandica,* **54**, 3–13.

PRICE, R. M. AND BONETT, D. G. (2004). "An improved confidence interval for a linear function of binomial proportions", *Computational Statistics & Data Analysis,* **45**, 449–456.

RAATS, V. M. AND MOORS, J. J. A. (2003). "Double-checking auditors: a Bayesian approach", *The Statistician,* **52**, 351–365.

TENENBEIN, A. (1970). "A double sampling scheme for estimating from binomial data with misclassifications", *Journal of American Statistical Association,* **65**, 1350–1361.

TENENBEIN, A. (1971). "A double sampling scheme for estimating from binomial data with misclassifications: sample size determination", *Biometrics,* **27**, 935–944.

TENENBEIN, A. (1972). "A double sampling scheme for estimating from misclassified multinomial data with applications to sampling inspection", *Technometrics,* **14**, 187–202.

YORK, J., MADIGAN, D., HEUCH, I. AND LIE, R. T. (1995). "Birth defects registered by double sampling: a Bayesian approach incorporating covariates and model uncertainty", *Applied Statistics,* **44**, 227–242.