Evaluating Alternative One-Sided Coverage Intervals for an Extreme Binomial Proportion

Yan K. Liu, Statistics of Income Division, IRS
Phillip S. Kott, Research and Development Division, NASS

Abstract. The interval estimation of a binomial proportion is difficult, especially when the proportion is extreme (very small or very large). Most of the methods discussed in the literature implicitly assume simple random sampling. These interval-estimation methods are not immediately applicable to data derived from a complex sample design. Some recent papers have addressed this problem, proposing modifications for complex samples. Matters are further complicated when a one-sided coverage interval is desired. This paper provides an extensive review of existing methods for constructing coverage intervals for a binomial proportion under both simple random and complex sample designs. It also evaluates the empirical performances of different one-sided coverage intervals under both a simple random and a stratified random sample design.

Key words: coverage probability, effective sample size, stratified random sample.

1. Introduction

Because of the poor performance of the standard Wald method for constructing coverage (confidence) intervals of a binomial proportion, the literature contains a series of modifications, alternative methods, and comparisons for a two-sided coverage interval under a simple random sample design (Brown et al. 2001, Agresti and Coull 1998, Vollset 1993, Clopper and Pearson 1934). Some recent papers have addressed this problem under more complex sample designs (Feng 2006, Sukasih and Jang 2006, Kott et al. 2001, Korn and Graubard 1998).

Constructing empirically effective one-sided coverage intervals can be even more difficult than two-sided intervals. Cai (2004) and Hall (1981) used Edgeworth expansion to develop one-sided coverage intervals under a simple a random sample. Kott and Liu (2007) modified Hall's method and extended it to handle data from a complex sample design with a particular emphasis on stratified (simple) random sampling.

We are interested here in constructing one-sided coverage intervals for proportions that are either very small (less than 20%) or very large (more than 80%). Section 2 provides an extensive list of coverage-interval methods under simple random sampling and then compares them. Section 3 looks at interval methods modified to handle complex sample data and evaluates their performances under stratified random sampling. Section 4 contains a brief discussion or our results.

2. Interval Estimation Methods Under a Simple Random Sample

Let X follow a binomial distribution with parameters n and p. The parameter p is also called the binomial proportion. In the survey sampling setting, n is the sample size of a simple random sample. Let k a sampled element and x_k be either 0 or 1. Assuming that x_k follows the Bernoulli distribution with parameter p, the estimator for p from the sample is $\hat{p} = x/n$, where $x = \sum^n x_k$.

This section contains a summary of many of the interval-construction methods under simple random sampling that have appeared in the literature. All the methods assume that the population size is large enough to ignore finite population correction. The symbol z is used to denote the z-score of a standard normal distribution associated with the one-sided coverage intervals of interest. For 95% coverage intervals, the z-score is 1.645.

2.1 The Methods

Standard Wald interval

This is the best known and most commonly used interval. It is based on the limiting distribution (as n grows arbitrarily large): $(\hat{p}-p)\big/\sqrt{v(\hat{p})} \to N(0,1)$, where $v(\hat{p}) = \hat{p}(1-\hat{p})/(n-1)$. The lower and upper bounds are

$$L_S = \hat{p} - z\sqrt{\hat{p}\left(1-\hat{p}\right)/\left(n-1\right)},$$
$$U_S = \hat{p} + z\sqrt{\hat{p}\left(1-\hat{p}\right)/\left(n-1\right)}. \qquad (1)$$

That is to say, the two one-sided Wald intervals for p are $p \geq L_S$, and $p \leq U_S$.

Wilson (Score) Interval

Instead of using the variance estimator for $\hat{p}$, this interval employs the true variance $V(\hat{p}) = p(1-p)/n$. It is based on the limit: $(\hat{p}-p)\big/\sqrt{V(\hat{p})} \to N(0,1)$. The lower and upper bounds are

$$L_W = \tilde{p} - \frac{z\sqrt{n}}{n+z^2}\sqrt{\hat{p}(1-\hat{p}) + \frac{z^2}{4n}}\ ,$$
$$U_W = \tilde{p} + \frac{z\sqrt{n}}{n+z^2}\sqrt{\hat{p}(1-\hat{p}) + \frac{z^2}{4n}}\ , \tag{2}$$

where $\tilde{p} = \dfrac{\hat{p} + z^2/2n}{1 + z^2/n}$.

Logit Interval

A logistic transformation, $\hat{\lambda} = \log\left[\hat{p}/(1-\hat{p})\right]$ stabilizes the variance of $\hat{p}$. The logit interval is based on the limit: $(\hat{\lambda}-\lambda)/\sqrt{v(\hat{\lambda})} \to N(0,1)$, where $v(\hat{\lambda}) = 1/\left[n\hat{p}(1-\hat{p})\right]$. The lower and upper bounds are

$$L_L = \frac{e^{\lambda_L}}{1+e^{\lambda_L}},\ \text{ where } \lambda_L = \hat{\lambda} - z\sqrt{v(\hat{\lambda})}\ ,$$
$$U_L = \frac{e^{\lambda_U}}{1+e^{\lambda_U}},\ \text{ where } \lambda_U = \hat{\lambda} + z\sqrt{v(\hat{\lambda})}\ . \tag{3}$$

Arcsine(root) Interval

Another transformation-stabilizing variance is the arcsine(root) transformation, $\delta = \arcsin(\sqrt{p})$. The interval for δ is based on the limit: $(\hat{\delta}-\delta)/\sqrt{v(\hat{\delta})} \to N(0,1)$, where $\hat{\delta} = \arcsin(\sqrt{\hat{p}})$ and $v(\hat{\delta}) = 1/(4n)$. This results in these lower and upper bounds for p:

$$L_A = \sin^2(\delta_L) = \sin^2\left[\arcsin(\hat{\delta}) - z/(2\sqrt{n})\right],$$
$$U_A = \sin^2(\delta_L) = \sin^2\left[\arcsin(\hat{\delta}) + z/(2\sqrt{n})\right]. \tag{4}$$

Jeffrey's Interval

The Bayesian Posterior interval under a Jeffrey's prior of the Beta distribution $Beta(1/2,\ 1/2)$ is

$$L_J = Beta(\alpha/2; x+1/2, n-x+1/2)\ ,$$
$$U_J = Beta(1-\alpha/2; x+1/2, n-x+1/2)\ . \tag{5}$$

Clopper-Pearson Exact Interval

This interval is based on inverting the equal-tailed binomial tests of the null hypothesis $H_0 : p = p_0$ against the alternative hypothesis $H_1 : p \neq p_0$. The lower and upper bounds can be obtained by solving the polynomial equations:

$$L_{CP} = \left\{ p : \sum_{t=0}^{x-1} \binom{n}{t} p^t (1-p)^{n-t} = 1 - \alpha/2 \right\}$$
$$U_{CP} = \left\{ p : \sum_{t=0}^{x} \binom{n}{t} p^t (1-p)^{n-t} = \alpha/2 \right\}. \tag{6}$$

They can be expressed in terms of Beta distribution as

$$L_{CP} = \text{Beta}(\alpha/2\,; x, n-x+1)\ ,$$
$$U_{CP} = \text{Beta}(1\text{-}\alpha/2\,; x+1, n-x)\ . \tag{7}$$

Mid-P Clopper-Pearson Interval

One way to reduce the perceived over-conservativeness of the Clopper-Pearson method obtains by solving the polynomial equations:

$$p_L = \left\{ p : \frac{1}{2}\binom{n}{x} p^x (1-p)^{n-x} + \sum_{t=0}^{x-1} \binom{n}{t} p^t (1-p)^{n-t} = 1 - \frac{\alpha}{2} \right\}$$
$$p_U = \left\{ p : \frac{1}{2}\binom{n}{x} p^x (1-p)^{n-x} + \sum_{t=0}^{x-1} \binom{n}{t} p^t (1-p)^{n-t} = \frac{\alpha}{2} \right\}.$$

The interval can be expressed in terms of Beta distribution as

$$L_{MP} = \frac{1}{2}\left\{ \text{Beta}\left(\frac{\alpha}{2}; x, n-x+1\right) + \text{Beta}\left(\frac{\alpha}{2}; x+1, n-x\right) \right\}$$

$$U_{MP} = \frac{1}{2}\left\{ \text{Beta}\left(1-\frac{\alpha}{2}; X, n-X+1\right) + \text{Beta}\left(1-\frac{\alpha}{2}; X+1, n-X\right) \right\} \tag{8}$$

Poisson Interval

When n is large and p is close to 0, the binomial distribution Bin(n, p) can be approximated by Poisson distribution $P(X = x) = \lambda^x e^{-\lambda}/x!$, where $\lambda = np$. The lower and upper bounds for p are

$$L_P = \chi^2_{2x,\alpha/2}\big/(2n)\ ,$$
$$U_P = \chi^2_{2(x+1),1-\alpha/2}\big/(2n)\ . \tag{9}$$

The nine methods described above can be used to construct both two-sided and one-sided intervals.

Unfortunately, an effective two-sided-interval method may not work as well in constructing a one-sided interval. This is because a two-sided interval can have compensating one-sided errors due to $\hat{p}$ being asymmetric. The following methods are based on an Edgeworth expansion that explicitly adjusts for the skewness in $\hat{p}$.

Hall Interval

The bounds for this interval translate the Wald bounds in equation (1) towards ½. They are

$$L_{KL} = \hat{p} + \delta - z\sqrt{v(\hat{p})}$$
$$U_{KL} = \hat{p} + \delta + z\sqrt{v(\hat{p})}\,, \qquad (10)$$

where $v(\hat{p}) = \dfrac{\hat{p}(1-\hat{p})}{n-1}$ and $\delta = \left(\dfrac{z^2}{3} + \dfrac{1}{6}\right)\dfrac{(1-2\hat{p})}{n}$.

The translation term, δ, is $O_P(1/n)$. Terms of smaller asymptotic order have been dropped. Hall (1982) has n in the denominator of $v(\hat{p})$ rather than n–1. This difference has no practical consequence when $n \geq 30$.

Cai Interval

Cai (2004) went further than Hall in correcting for the skewness in $\hat{p}$ by keeping $O_P(1/n^2)$ terms producing the bounds:

$$L_{Cai} = \breve{p} - \frac{z}{\sqrt{n}}\sqrt{\hat{p}(1-\hat{p}) + \frac{\gamma_1 \hat{p}(1-\hat{p}) + \gamma_2}{n}}\,,$$
$$U_{Cai} = \breve{p} + \frac{z}{\sqrt{n}}\sqrt{\hat{p}(1-\hat{p}) + \frac{\gamma_1 \hat{p}(1-\hat{p}) + \gamma_2}{n}}\,, \qquad (11)$$

where $\breve{p} = \dfrac{\hat{p} + \eta / n}{1 + 2\eta / n}$, $\eta = \dfrac{z^2}{3} + \dfrac{1}{6}$,

$\gamma_1 = -\dfrac{13}{18}z^2 - \dfrac{17}{18}$ and $\gamma_2 = \dfrac{1}{18}z^2 + \dfrac{7}{36}$.

Kott-Liu Interval

Under simple random sampling, Kott and Liu proposed a slight modification of the Hall interval that better handles samples with small $\hat{p}(1-\hat{p})$ values:

$$L_{KL} = \hat{p} + \delta - \sqrt{z^2 v(\hat{p}) + \delta^2}$$
$$U_{KL} = \hat{p} + \delta + \sqrt{z^2 v(\hat{p}) + \delta^2}\,, \qquad (12)$$

where $v(\hat{p})$ and δ are unchanged. This method will be described further in the following section.

Other Intervals

There are also various continuity-correction approaches that are not included in this paper. Two other methods not treated here are the Wilson-logit and likelihood-ratio interval. These methods employ an iteration algorithm to obtain the interval end-points.

2.2 Comparison of One-Sided Intervals Under Simple Random Sampling

In this subsection, the methods defined in equations (1) through (12) are used to construct one-sided 95% coverage intervals. They are then compared in terms of their coverage probabilities and the average distances from their endpoints to the true value of p.

The *coverage probability* for the given p and n is defined as the probability of p falling within the coverage interval CI, that is,

$$P(p \in CI) = \sum_{x=0}^{n} I(x)P(x)\,,$$

where $CI = \begin{cases} (0, L), & \text{for lower bound} \\ (\text{U}, 1) & \text{for upper bound} \end{cases}$

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad 0 < p < 1\,,$$

and $I(x) = \begin{cases} 1, \text{ if } p \in CI \\ 0, \text{ if } p \notin CI \end{cases}$.

The *average distance* for the given p and n is defined here as the mean of the absolute distance of lower or upper bound from the true value of p, that is,

$$AD = \sum_{x=0}^{n} D(x)P(x)\,,$$

where $D(x) = \begin{cases} |L(x) - p|, & \text{for the lower bound} \\ |U(x) - p|, & \text{for the upper bound} \end{cases}$

It is well known that the coverage intervals of a binomial proportion behave irregularly (Brown, Cai and DasGupta, 2001 & 2002). A coverage interval will perform differently for different sample sizes and different values of p. We are interested in the setting where the sample size n is reasonably large – at least 30 – and the value of p is either small or large. Therefore, we evaluate sample size of 30, 60 and 120 and focus on the comparison for the value of p in the range of (0, 0.20) and (0.80, 1). We also modify the intervals at x =0, 1. First, we force the lower bound to be 0 at x =0 and 1 at x =1. Second, when the lower bound or upper bound is not defined at x =0, 1 for some methods (Wald, Logit and Mid-P), we replace them with the Clopper-Pearson method.

Except for the Poisson, the coverage probabilities and average distances for all the methods are symmetric or very nearly so in the range $0 \le p \le 1$. Consequently, conclusions drawn about lower bounds for $p<0.2$ also apply to upper bounds for $p>0.8$, and conclusions about lower bounds for $p>0.8$ apply to upper bounds for $p<0.2$. Because of this, we only calculate coverage probabilities and average distances for lower bounds. These values are calculated at $p = 0.001, 0.002, 0.003, \ldots, 0.998, 0.999$.

Due to the space limitation, the plots are not displayed here. The following conclusions about the coverage probabilities of the methods can be drawn from them:

- Wald and Arcsine are systematically biased, sometimes in one direction sometimes in the other.
- Poisson is overly conservative, that is, has coverages well above the nominal rate (95%). It should not beviewed as a serious competitor to the other methods.
- Clopper-Pearson always has at least the nominal coverage, but often over-covers.
- Wilson and Logit are systematically biased in the opposite direction of Wald and to a lesser degree. They tend to under-cover for small *p* and over-cover for large *p*. The overage-coverage for Wilson near $p=1$ is not as pronounced as for Clopper-Pearson.
- Jeffrey and Hall have large downward spikes (under-coverages) near the two boundaries.
- Mid-P has large downward spikes near $p=0$, but performs well for large *p*.
- Kott-Liu and Cai provide reasonably coverages everywhere with Kott-Liu having slightly smaller oscillations near $p=1$.

These conclusions, which obtain when m =30, 60 or 120, are summarized in Table 1.

We plot the average distances of lower bounds versus the values of p for the 'Best Pick' methods and for the conservative Clopper-Pearson. In general, the average distance is longer when the coverage probability is larger. Due to the space limit, the plots are not presented here. Clopper-Pearson has a much longer average distance than the other methods, not surprising since it tends to be conservative. For small *p*, Kott-Liu and Cai behave very similarly. For large *p*, Kott-Liu tends to be slightly longer than Cai. Wilson is longer than both Kott-Liu and Cai. Mid-P becomes longer than Kott-Liu and Cai when *p* gets near 1 but not before.

In summary, Kott-Liu and Cai are the best in terms of having coverages always reasonably close to the nominal. Clopper-Pearson, never under-covers, which some find a desirable characteristic, but has longer average distances.

Table 1. Comparison in terms of Lower-Limit Coverage Probabilities

Method	p<0.2	p>0.8
Wald	Systematic biased	
Arcsine		
Poisson	Over-conservative	Not applicable
Clopper-Pearson	Conservative	
Wilson	Under-coverage, Large Downward spikes near *p*=0	Conservative, not as much as Clopper-Pearson
Logit	Under-coverage, Large Downward spikes near *p*=0	Conservative, as much as Clopper-Pearson
Jeffrey	Have large downward spikes	Have large downward spikes
Hall		
Mid-P	Good coverage, except for *p* near 0 (large spikes near *p*=0)	Good coverage
Cai	Good coverage	Good coverage
Kott-Liu	Good coverage, slightly smaller oscillations than Cai	Good coverage, slightly smaller oscillations than Cai
Best Pick	**Kott-Liu, Cai**	**Kott-Liu, Cai, Mid-P, Wilson (conservative)**

3. Interval Construction Methods Under Stratified Random Sampling

Let s denote elements of the whole sample, k (again) denote an element, and w_k the weight of element k. Let x_k be either 0 or 1. The estimated proportion is then $\hat{p} = \sum_s x_k w_k \Big/ \sum_s w_k$.

3.1 The Methods

The most common way of extending interval-construction methods to handle sample data from a complex design is by replacing the sample size n with the effective sample size n^* and replacing x with $x^*=n^*\hat{p}$. When $v(\hat{p})>0$, where $v(\hat{p})$ is the estimated variance of $\hat{p}$ under the complex sample design, the effective sample size n^* can be defined as

$$n^* = \frac{n}{DEFF(\hat{p})} = \frac{\hat{p}(1-\hat{p})}{v(\hat{p})} \tag{13}$$

(alternatively, n^* can be defined as 1 plus the left-hand side of equation (13); the distinction is usually trivial when $n \geq 30$).

This *ad hoc* procedure was used and discussed in Kott and Carr (1997) for modifying the Wilson interval and in Korn and Graubard (1998) for modifying the Clopper-Pearson interval. Feng (2006) treated a few other intervals with this procedure.

We focus in this section on an empirical evaluation of the alternative methods under stratified random sampling. We apply the effective sample size procedure to all the methods from Section 2 except the Kott-Liu, which was designed especially to handle data from stratified random samples. We follow Korn and Graubard and set $n^*=n$ when $v(\hat{p})=0$.

Let $W_h = N_h/N$ for a stratified random sample with H strata. The estimated overall proportion is $\hat{p}=\sum^H W_h\hat{p}_h$, where $\hat{p}_h$ is the observed stratum proportion of stratum h.

Adapting the Edgeworth expansions in Hall and Cai, Kott and Liu (2007) actually discuss three different coverage intervals for data from a stratified random sample.

<u>*Basic Kott-Liu Interval*</u>

$$L_{KL1} = \hat{p}+\delta_1-\sqrt{z^2 v_1(\hat{p})+\delta_1{}^2}$$
$$U_{KL1} = \hat{p}+\delta_1+\sqrt{z^2 v_1(\hat{p})+\delta_1{}^2}\ , \tag{14}$$

where $v_1(\hat{p}) = \sum_h W_h^2\hat{p}_h(1-\hat{p}_h)/(n_h-1)$

and

$$\delta_1 = \left(\frac{z^2}{3}+\frac{1}{6}\right)\frac{\sum_h W_h^3\hat{p}_h(1-\hat{p}_h)(1-2\hat{p}_h)\Big/\left[(n_h-1)(n_h-2)\right]}{\sum_h W_h^2\hat{p}_h(1-\hat{p}_h)\Big/(n_h-1)}$$

The variance of $\hat{p}$ is not a simple function of the true p and n under stratified random sampling as it is under simple random sampling. As a result, V($\hat{p}$) must be estimated from the sample. This estimation has its own random error, which cannot be completely eliminated from the Edgeworth expansion (moreover, keeping $O_P(1/n^2)$ terms, like Cai does, becomes impossible). The following interval attempts to account for that additional source of error.

<u>*DF-adjusted Kott-Liu Interval*</u>

Replacing the z-score in equation (14) with a *t*-score from a Student *t* distribution can reduce the downward spikes when *p* is near 0 or 1. A *t*-distribution needs a degrees-of-freedom calculation. Kott and Liu discuss a number of ways of estimating the *effective degree of freedom.* When each stratum has at least 10 observations, a nearly unbiased estimator for this quantity is

$$df_1 = \frac{2a_1^2}{a_3 - a_2^2/a_1}\ ,$$

where $a_1 = \sum_h W_h^2\hat{p}_h(1-\hat{p}_h)/n_h$,

$$a_2 = \sum_h W_h^3\hat{p}_h(1-\hat{p}_h)(1-2\hat{p}_h)/n_h^2\ ,$$

$$a_3 = \sum_h W_h^4\hat{p}_h(1-\hat{p}_h)(1-2\hat{p}_h)^2/n_h^3\ .$$

An asymptotically biased, but more stable, effective-degrees-of-freedom estimator treats the p_h as if they were equal:

$$df_2 = \frac{2\left(\sum_h W_h^2/n_h\right)^2\hat{p}(1-\hat{p})}{\left\{\sum_h\frac{W_h^4}{n_h^3}-\left(\sum_h W_h^3/n_h^2\right)^2\Big/\sum_h\frac{W_h^2}{n_h}\right\}(1-2\hat{p})^2}$$

A slightly conservative policy (justified by observation) sets the estimated effective degrees of freedom at $df = Min(df_1, df_2)$ and uses $t(df, 1-\alpha)$ in place of z in the lower and upper bounds defined in equation (13).

Kott-Liu iid Interval

If an independent and identically distributed (*iid)* Bernoulli model is assumed, then a different way to generalize equation (12) is with

$$L_{KL2} = \hat{p} + \delta_2 - \sqrt{z^2 v_2(\hat{p}) + \delta_2{}^2}$$
$$U_{KL2} = \hat{p} + \delta_2 + \sqrt{z^2 v_2(\hat{p}) + \delta_2{}^2}\ , \qquad (15)$$

where $v_2(\hat{p}) = \sum_h W_h^2 \hat{p}(1-\hat{p})/n_h$

and $\delta_2 = \left(\frac{1-z^2}{6} \frac{\sum_h W_h^3/n_h^2}{\sum_h W_h^2/n_h} + \frac{z^2}{2} \sum_h \frac{W_h^2}{n_h} \right)(1-2\hat{p})$

Since both the basic and DF-adusted Kott-Liu intervals are undefined when $\hat{p}$=0 or 1, Kott and Liu suggest using their *iid* method in equation (15) in this situation. In fact, when $\hat{p}$ is near 0 or 1, it makes sense to use the *iid* method as the proportions cannot vary very much across the strata.

3.2 Comparison of One-Sided Intervals under Stratified Random Sampling

All the methods described in the text are compared under a stratified random sampling design using simulations. A population of 6,000 is divided into 3 equal strata, that is, N_h=2,000, $h = 1,2,3$. The overall proportion *p* takes the values of 0.001, 0.002, 0.003, ..., 0.998, 0.999. The settings for the four stratum sample size allocations and comparative values for the p_h are shown in Table 2.

Table 2. Simulation Settings

Stratum Sample Size Allocation (n_1, n_2, n_3)	Allocation of Binomial Proportion (p_1, p_2, p_3)	
	(*p, p, p*)	(*p, p-pq, p+pq*)
10, 10, 10	A	E
15, 15, 15	B	F
10, 15, 20	C	G
20, 15, 10	D	H

The first sample size allocation, (10, 10, 10), has a total sample size of 30, our minimum. The other three each have a total sample size of 45 with the minimum stratum sample size being 10. One setting for the comparative stratum values of the p_h features proportional allocation, (*p, p, p*). The other, (*p-pq, p, p+pq*; where *q* = 1–*p*), in some sense maximizes the spread of the p_h while being symmetrical and keeping all p_h in the 0 to 1 range.

In the simulations, we first generate a finite population of 2,000 units in each stratum *h*, denoted as x_{hi} = 1, 2, ….., 2,000. We then draw 1,000 stratified random samples for each stratum sample size allocation. For each stratum proportion p_h , we set

$$y_{hi} = \begin{cases} 1, & \text{if } x_{hi} < 2,000 p_h \\ 0, & \text{otherwise} \end{cases} .$$

The weighted estimate for the proportion of *y*=1 is calculated for each value of *p* and for each sample. The coverage intervals are constructed using the methods described earlier in the text with the coverage probabilities and the average distances calculated from the 1,000 samples for each *p*.

Analogously with the simple random sample sampling case, only the simulation results for a lower bound need be considered. Due to the space limitation, we only display figures for the simulation setting A. Similar conclusions hold for other settings with larger sample sizes and proportional allocation leading to better coverage probabilities across virtually all the methods.

Figure 1 (on the last page) plots the coverage probabilities versus values of *p* for the sample size setting (10, 10, 10) and the p_h setting (*p, p, p*). As shown in Figure 1,

- Wald and Arcsine have large biases and large oscillations in the coverage.
- Poisson has large coverage probabilities, very close to 1 when *p*>0.5.
- Clopper-Pearson is conservative with coverage probabilities almost always above the nominal level.
- When *p* is in mid-range, say from 0.2 to 0.8, There are many good methods such as Jeffrey, Mid-P, Cai, Hall and Kott-Liu.
- When *p* is near 0, Cai and Kott-Liu methods perform reasonable well and better than the others.
- When *p* is near 1, Kott-Liu methods work fairly well. The basic and DF-adjusted versions are virtually identical. Estimating the effective degrees of freedom has little to no effect.
- When *p* is near 1, Cai and Mid-P are also reasonable candidates, with the Mid-P getting more conservative than the others as *p* grows closer to 1. Like the Kott-Liu, these become extremely conservative very near 1.

Figure 2 shows the average distances of lower bounds for four methods. For *p* small, Mid-P, Cai and DF-adjusted Kott-Liu methods have similar average distances, much shorter than Clopper-Pearson. For *p* large, but not near 1, Mid-P, Kott-Liu and Cai are close, and much shorter than Clopper-Pearson. When *p* gets near 1, Mid-P gets longer than Cai and Kott-Liu. The average distance of the DF-adjusted Kott-Liu is slightly longer than Cai, while DF-adjusted Kott-Liu has a slightly superior coverage.

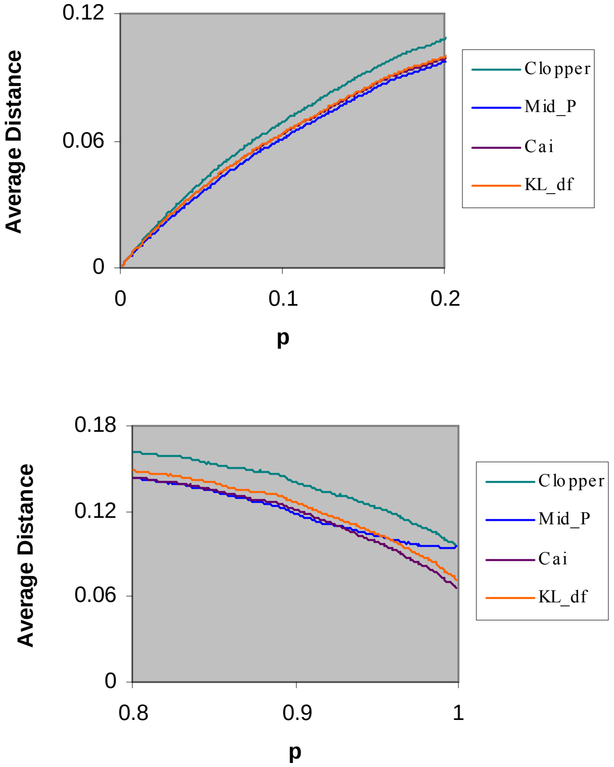

Figure 2. Average Distances of Lower Bounds at 95% Nominal Level for Simulation Setting A

4. Discussion

After reviewing much of the literature on constructing one-sided coverage intervals under simple random sampling, we conducted our own empirical evaluation and found that, among the methods considered, the Cai and Kott-Liu had coverages closest to nominal. We also confirmed that the Clopper-Pearson method always provided at least the nominal coverage, which many find reassuring.

When we turned to stratified random sampling. Applying the effective-sample-size technique to the Clopper-Pearson (Korn-Graubard method) was still conservative with coverage probabilities almost always over the nominal level except when the sample size allocation is disproportional and p is near 1 for lower bound and near 0 for upper bound. The Kott-Liu methods appeared slightly superior to the others, with the *iid* version having problems (not shown) when the stratum proportions are unequal. Adjusting the basic Kott-Liu method by its effective degrees of freedom did little in our simulations except under certain settings (not shown). We also looked at more simulations for settings not listed in Table 2 and found that the proportional allocation of sample size gives a much better coverage probability than a disapportional allocation.

5. REFERENCE

Brown, L.D., Cai, T. and Dasgupta, A. (2001). Interval Estimation for a Binomial Proportion. *Statistical Science*, **16**, 101-133.

Brown, L.D., Cai, T. and Dasgupta, A. (2002). Confidence Intervals for a Binomial Proportion and Asymptotic Expansions. *The Annals of Statistics*, Vol 30, No. 1, 160-201.

Cai, T. (2004). One-sided confidence intervals in discrete distributions. *Journal of Statistical Planning and Inference*, Vol. 131, No. 1. pp. 63-88.

Feng, X. (2006). Confidence intervals for proportions with focus on the US National Health and Nutrition Examination Survey. *Master Thesis. Simon Fraiser University.*

Hall, P. (1982). Improving the normal approximation when constructing one-sided confidence intervals for binomial or Poisson parameters. *Biometrika,* 69(3):647-652.

Korn, E. L. and Graubard B. I. (1998). Confidence Intervals for Proportions With Small Expected Number of Positive Counts Estimated From Survey Data. *Survey Methodology*, Vol. 24, No. 2, pp. 193-201.

Kott, P. S. and Liu, Y. K. (2007). One-sided coverage intervals for a Proportion Estimated from a Stratified Simple Random Sample (in preparation).

Kott, P. S. and Carr, D. A. (1997). Developing an Estimation Strategy for a Pesticide Data Program. *Journal of Official Statistics*, 13, 367-383.

Newcombe, R. G. (1998). Two-sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods. *Statistics in Medicine*, 17, 857-872.

Vollset, S. E. (1993). Confidence Intervals for a Binomial Proportion. *Statistics in Medicine*, 12, 809-827.

The authors express a special note of thanks to Kevin Cecco at IRS for his encouragement and providing some reference papers at the beginning of this project.

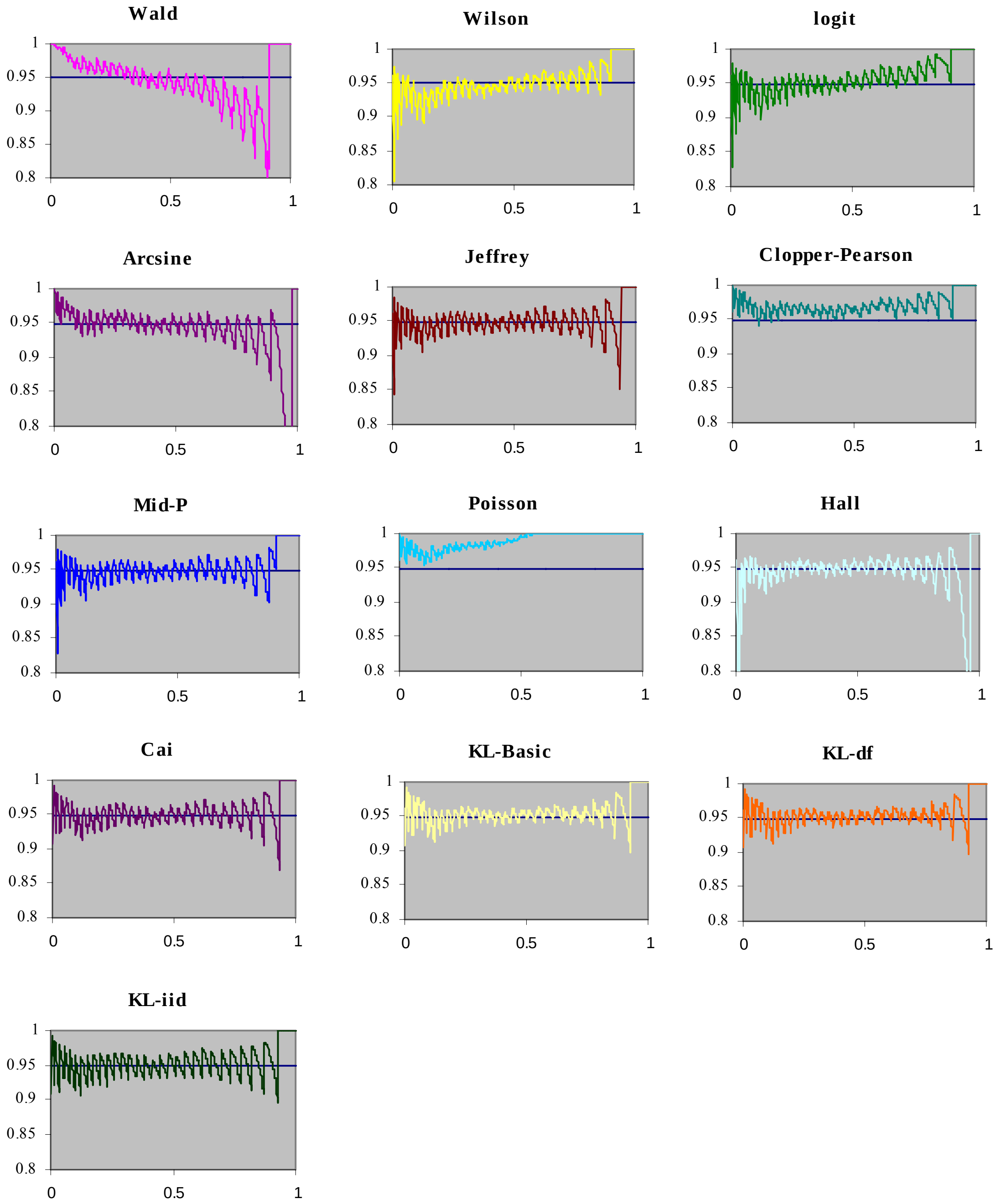

Figure 1. Coverage Probabilities of Lower Bounds at 95% Nominal Level for Simulation Setting A