# Two Papers on One-sided Coverage Intervals for a Proportion

Phillip S. Kott and Yan K. Liu

These two papers are the result of research conducted by the authors on methods for constructed one-sided coverage intervals based on relatively small simple random or stratified simple random samples. The first paper lays out the theory for a new method. The second paper evaluates that method and a number of alternatives.

KEY WORDS: Wald intervals; Edgeworth Expansion; Sampling Weight; Asymptotic; Skewness.

Phillip S. Kott is Chief Research Statistician, Research and Development Division, National Agricultural Statistics Service, 3251 Old Lee Highway, Room 305, Fairfax, Virginia, 22030. Yan K. Liu is a mathematical statistician at the Statistics of Income Division, Internal Revenue Service.

# One-Sided Coverage Intervals for a Proportion Estimated from a Stratified Simple Random Sample

Phillip S. Kott and Yan K. Liu

National Agricultural Statistics Service, Fairfax (VA) and

Internal Revenue Service, Washington (DC)

Using an Edgeworth expansion to speed up the asymptotics, we develop one-sided coverage intervals for a proportion based on a stratified simple random sample. To this end, we assume the population units are independent random variables with a common mean within each stratum. These stratum means, in turn, may either be free to vary or are assumed to be constant. The more general assumption is equivalent to a model-free randomization-based framework when finite population correction is ignored. Unlike when an Edgeworth expansion is used to construct one-sided intervals under simple random sampling, it is necessary to estimate the variance of the estimator for the population proportion when the stratum means are allowed to differ. As a result, there may be accuracy gains from replacing the Normal $z$-score in the Edgeworth expansion with a $t$-score.

**Keywords:** Edgeworth expansion, effective degrees of freedom, model.

Contact information: National Agricultural Statistics Service, 3251 Old Lee Highway, Fairfax, VA 22050 USA (pkott@nass.usda.gov)

## 1. Introduction

Suppose we have a population containing $N$ units separated into $H$ strata. Let $p_h$ be the proportion of units in stratum $h$ with a particular property. Formally, we can write $y_i = 1$ when unit $i$ has the property and $y_i = 0$ otherwise, so that $p_h = \sum y_i / N_h$, where the summation is over the $N_h$ units in stratum $h$.

Our goal is to estimate the overall proportion of the population with the property; that is, $p = \sum W_h p_h$, where $W_h = N_h / N$, and the summations across the $H$ strata. Given a stratified simple random sample, an unbiased estimator for $p$ is $\hat{p} = \sum W_h \hat{p}_h$, where $\hat{p}_h$ is the proportion of the $n_h$ sampled units in stratum $h$ with the property. The overall sample size is denoted $n = \sum n_h$.

For analyzing the statistical properties of $\hat{p}$, we can use either model-free randomization-based inference or assume a *general model* in which each unit in the population is independent with probability $p_h$ of having the property depending on its stratum membership. The latter approach allows us to ignore finite population correction, which we will do from now on. Although based on a model this approach is completely analogous to randomization-based inference under stratified simple random sampling *with* replacement.

We will also consider an alternative model-based framework in which each unit is independent and identically distributed. Under this *iid model*, the $p_h$ are assumed to be equal (at least) for variance-estimation and coverage-interval construction purposes. Note, however, that the estimator $\hat{p} = \sum W_h \hat{p}_h$ is computed as if there is a possibility that the $p_h$ might not be equal. In other words, $\hat{p}$ is (or, more correctly, can be) a weighted averages of the sampled $y_i$.

A one-sided $\alpha$-percent Wald coverage interval for $p$ has the form

$$p \le \hat{p} + \Phi^{-1}(\alpha)\sqrt{v(\hat{p})} \quad or$$
$$p \ge \hat{p} - \Phi^{-1}(\alpha)\sqrt{v(\hat{p})}$$

where $v(\hat{p})$ is an estimator for $V(\hat{p}) = \sum W_h^2 p_h(1 - p_h)/n_h$, the variance of $\hat{p}$, and $\Phi(.)$ is the cumulative distribution function (*cdf*) of a standard normal distribution. One can show that if the sample size is large enough, then both inequalities will hold for roughly $\alpha$-percent of the samples that could be drawn using a fixed set of $n_h$.

In practice, the sample size is often not nearly large enough for a one-sided Wald intervals to contain ("cover") $p$ with the frequency suggested by theory when either $p$ or $1-p$ is small. As a consequence of this, a host of alternative coverage intervals for $p$ have been proposed in the literature. Most focus on simple random samples from an *iid* population, but a few treat samples from more complex designs. See Lui and Kott (2007) for descriptions and empirical evaluations of many of them. The emphasis there, as here, is on one-sided coverage intervals. Some techniques known to produce reasonably effective two-sided intervals, such as Wilson's score method, can fail when one-sided-interval construction is the goal. See Cai (2004).

When all $n_h \ge 3$, we will propose the following pair of one-sided $\alpha$-percent coverage intervals for $p$ in a stratified population under the general model:

$$p \le \hat{p} + \delta + \sqrt{z^2 v(\hat{p}) + \delta^2}, \text{ and}$$
$$p \ge \hat{p} + \delta - \sqrt{z^2 v(\hat{p}) + \delta^2} \tag{1}$$

where $v(\hat{p}) = v_1(\hat{p}) = \sum_{h=1}^{H} \frac{W_h^2}{n_h - 1} \hat{p}_h (1 - \hat{p}_h),$ (2.1)

$$\delta = \delta_1 = \left(\frac{1}{6} + \frac{z^2}{3}\right) \frac{\sum_{h=1}^{H} \frac{W_h^3}{(n_h - 1)(n_h - 2)} \hat{p}_h (1 - \hat{p}_h)(1 - 2\hat{p}_h)}{\sum_{h=1}^{H} \frac{W_h^2}{n_h - 1} \hat{p}_h (1 - \hat{p}_h)}, \tag{2.2}$$

and $z = \Phi^{-1}(\alpha)$. At first, $\Phi(.)$ will (again) be the *cdf* of a standard normal distribution. Later, we suggest replacing it by the *cdf* of a Student's *t* distribution with specified degrees of freedom.

An alternative pair of one-sided intervals can be developed for $p$ when the *iid* model holds. They have the same form as in equation (1), but equation (2) is replaced with

$$v(\hat{p}) = v_2(\hat{p}) = \left(\sum_{h=1}^{H} \frac{W_h^2}{n_h}\right) \hat{p}(1 - \hat{p}), \tag{3.1}$$

and $$\delta = \delta_2 = \left( \frac{1 - z^2}{6} \frac{\sum_{h=1}^{H} \frac{W_h^3}{n_h^2}}{\sum_{h=1}^{H} \frac{W_h^2}{n_h}} + \frac{z^2}{2} \sum_{h=1}^{H} \frac{W_h^2}{n_h} \right) (1 - 2\hat{p}). \tag{3.2}$$

In practical application, when $\hat{p}_1 = \hat{p}_2 = ... = \hat{p}_H = \hat{p} = 0$, we are effectively forced to assume the *iid* model and use equation (3) instead of equation (2). Equation (3) can also be used when not every stratum has at least three sampled units. It must be remembered, however, that its theoretical appropriateness depends on what can be a dubious the *iid* assumption.

Like their Wald counterparts, the coverage intervals in equation (1) depend on the sample being "sufficiently large." The insertion into equation (1) of a non-zero value for the $\delta$, whether defined in equation (2.2) or (3.2), effectively speeds up the asymptotics. The observant reader will note that when collapsed into the single-stratum environment where $H = 1$, equations (2) and (3) do not quite coincide: $v_1(\hat{p})$ and $v_2(\hat{p})$ differ by an $O_P(1/n^2)$ term as do $\delta_1$ and $\delta_2$. These differences are deemed asymptotically ignorable here.

Equation (1) modifies and generalizes coverage intervals introduced by Hall (1982) for a proportion of an unstratified *iid* population. In Hall's original formulation there is no $\delta^2$ under the radical. Our variation, which otherwise like Hall's drops $O_P(1/n^{3/2})$ terms, assures that the lower-tail interval $p \geq \hat{p} + \delta - \sqrt{z^2 v(\hat{p}) + \delta^2}$ contains zero when $\hat{p} = 0$ (and the upper-tail interval contains one when $\hat{p} = 1$).

Section 2 begins the derivation of our proposed coverage intervals by introducing an Edgeworth expansion for $\hat{p}$. Section 3 completes the derivation under the general model, while Section 4 treats the *iid* model. A generally favorable empirical evaluation of a particular stratified sampling design in Section 5 prompts a discussion of the limitations of our proposals, especially when constructing a lower-bound for $p$ when $\hat{p} = 0$ and an upper bound when $\hat{p} = 1$. Finally, we briefly discuss two-sided coverage intervals and potential extensions of our methodology in Section 6.

## 2. The Edgeworth Expansion

The derivation to follow parallels Cai (2004) more closely than Hall. Unlike Cai, however, we will be satisfied dropping $O_P(1/n^2)$ trems. To do otherwise seems to us an impossible task under the general model in a stratified setting because the variance of $\hat{p}$ cannot be expressed as an explicit function of $p$ and the $n_h$.

We begin with an Edgeworth expansion for $\hat{p}$:

$$\Pr\left( \frac{\hat{p} - p}{\sqrt{V(\hat{p})}} \leq z \right) = \Phi(z) + (1/6)(1 - z^2)\varphi(z)\tau + O(1/n), \tag{4}$$

where $\varphi(z)$ is the probability density function (*pdf*) of the standard normal distribution, and

$$\tau = \frac{E[(\hat{p} - p)^3]}{\left\{E[(\hat{p} - p)^2]\right\}^{3/2}} = \frac{M_3}{\left[V(\hat{p})\right]^{3/2}} = \frac{\sum_{h=1}^{H} \frac{W_h^3}{n_h^2} p_h(1 - p_h)(1 - 2p_h)}{\left[\sum_{h=1}^{H} \frac{W_h^2}{n_h} p_h(1 - p_h)\right]^{3/2}}$$

$$= \frac{O(1/n^2)}{\left[O(1/n)\right]^{3/2}} = O(1/n^{1/2})$$

under mild conditions on the $W_h$ and $p_h$ we assume to hold. Sufficient conditions are that all the $nW_h/n_n$ are bound from above by a finite $U$, while all the $p_h$ $(1 - p_h)$ are bound from below by a positive $L$.

Strictly speaking, Edgeworth expansions only apply for continuous distributions, while $\hat{p}$ has a discrete distribution since the number of population units in each stratum having the property of interest must be a whole number. Like Hall and Cai, we are ignoring the "oscillatory terms" of the *cdf*. As a result, one of our α-percent intervals will (at best) cover $p$ in α-percent of all possible samples on average across small ranges of potential values for *p* rather than covering *p* at in *at least* α-percent of all possible samples for *each p*. This is why we use the term "coverage interval" to describe the intervals in equation (1) rather than the more common "confidence interval." The latter suggests to us a very conservative requirement that coverage is at least as good as advertised no matter the true *p*. See Brown *et al*. (2001) and subsequent comments for a debate on this issue in the context of two-sided intervals for an unstratified *iid* population.

Letting

$$a = (1/6)(1-z^2)\frac{M_3}{\left[V(\hat{p})\right]^{3/2}} = O(1/n^{1/2}),$$

and employing the Taylor-series expansion, $\Phi(z-a) = \Phi(z) - a\varphi(z) + O(1/n)$, we can replace $z$ in equation (4) with $z - a$ and write:

$$\Pr\left(\frac{\hat{p}-p}{\sqrt{V(\hat{p})}} \leq z - (1/6)(1-z^2)\frac{M_3}{\left[V(\hat{p})\right]^{3/2}}\right) = \Phi(z) + O(1/n)$$

or equivalently

$$\Pr\left(\frac{\hat{p}-p}{\sqrt{V(\hat{p})}} \leq z - (1/6)(1-z^2)\frac{M_3}{\left[V(\hat{p})\right]^{3/2}} + O(1/n)\right) = \Phi(z).$$

This implies

$$\Pr\left(\hat{p}-p \le z\sqrt{V(\hat{p})}-(1/6)(1-z^2)\frac{M_3}{V(\hat{p})}+O(1/n^{3/2})\right)=\Phi(z),$$

$$\Pr\left(\left[\hat{p}-p+(1/6)(1-z^2)\frac{M_3}{V(\hat{p})}+O(1/n^{3/2})\right]^2 \le z^2V(\hat{p})\right)=\Phi(z),$$

and

$$\Pr\left(\left(p-\hat{p}\right)^2-(1/3)(1-z^2)\frac{M_3}{V(\hat{p})}\left(p-\hat{p}\right)+O(1/n^2)\le z^2V(\hat{p})\right)=\Phi(z).$$

To use this last equation in constructing coverage intervals, we will need (among other things) to estimate the unknown $V(\hat{p})$. In the unstratified framework explored by Hall and Cai, $V(\hat{p})$ is a known function of $p$. That need not be the case here. As a consequence, we follow Andersson and Nerman (2000) and replace $V(\hat{p})$, not with $v(\hat{p})$ as one might expect, but with the much more efficient *idealized variance estimator*:

$$v^*(\hat{p})=v(\hat{p})-B\left(\hat{p}-p\right), \tag{5}$$

where

$$B=\frac{Cov\left(v(\hat{p}),\hat{p}\right)}{V(\hat{p})}.$$

Unfortunately, $v^*(\hat{p})$, although having the minimum variance of an estimator for $V(\hat{p})$ in the form $v(\hat{p})-\lambda\left(\hat{p}-p\right)$, can still possess an error of probability order $1/n^{3/2}$.

Using $v^*(\hat{p})$ in equation (5) ultimately requires an estimator for $B$. At this point, we leave vague both how to compute that estimator and what to use for $v(\hat{p})$ (whether $v_1(\hat{p})$ from equation (2.1) or $v_2(\hat{p})$ from (3.1)).

Substituting $z^2 V(\hat{p})$ by $z^2 v^*(\hat{p})$ (in fact, $z^2 v^*(\hat{p})$ + $O_P(1/n^{3/2})$ ) and rearranging brings us to

$$\Pr\left( \left(p-\hat{p}\right)^2 - \left[\tfrac{1}{3}(1-z^2)\frac{M_3}{V(\hat{p})} + z^2 B\right]\left(p-\hat{p}\right) - z^2 v(\hat{p}) + O_P(1/n^{3/2}) \le 0 \right) = \Phi(z).$$

By solving the quadratic equation, we then have

$$\Pr\left( p-\hat{p} \le \frac{\tfrac{1}{3}(1-z^2)\dfrac{M_3}{V(\hat{p})} + z^2 B + \sqrt{\left[\tfrac{1}{3}(1-z^2)\dfrac{M_3}{V(\hat{p})} + z^2 B\right]^2 + 4z^2 v(\hat{P}) + O_P(1/n^{3/2})}}{2} \right) = \Phi(z)$$

or

$$\Pr\left( p-\hat{p} \ge \frac{\tfrac{1}{3}(1-z^2)\dfrac{M_3}{V(\hat{p})} + z^2 B - \sqrt{\left[\tfrac{1}{3}(1-z^2)\dfrac{M_3}{V(\hat{p})} + z^2 B\right]^2 + 4z^2 v(\hat{P}) + O_P(1/n^{3/2})}}{2} \right) = \Phi(z).$$

This means

$$\Pr\left( p \le \hat{p} + \tfrac{1}{6}(1-z^2)\frac{M_3}{V(\hat{p})} + \frac{z^2}{2}B + \sqrt{z^2 v(\hat{p}) + \left[\tfrac{1}{6}(1-z^2)\frac{M_3}{V(\hat{p})} + \frac{z^2}{2}B\right]^2 + O_P(1/n^{3/2})} \right) = \Phi(z)$$

or (6)

$$\Pr\left( p \ge \hat{p} + \tfrac{1}{6}(1-z^2)\frac{M_3}{V(\hat{p})} + \frac{z^2}{2}B - \sqrt{z^2 v(\hat{p}) + \left[\tfrac{1}{6}(1-z^2)\frac{M_3}{V(\hat{p})} + \frac{z^2}{2}B\right]^2 + O_P(1/n^{3/2})} \right) = \Phi(z).$$

## 3. The General Model

Under the general model, both $M_3 / V(\hat{p})$ and $B$ in equation (6) can be replaced with the same expression:

$$b_1 = \frac{m_{3(1)}}{v_1(\hat{p})} = \frac{\sum_{h=1}^{H} \frac{W_h^3}{(n_h - 1)(n_h - 2)} \hat{p}_h (1 - \hat{p}_h)(1 - 2\hat{p}_h)}{\sum_{h=1}^{H} \frac{W_h^2}{n_h - 1} \hat{p}_h (1 - \hat{p}_h)}. \tag{7}$$

(see, for example, Kott *et al.* 2001). This estimator is $O_P(1/n)$ and has an error of $O_p(1/n^{3/2})$, which is asymptotically dominated by other terms when plugged, along with $v_1(\hat{p})$, into equation (6). By dropping the unspecified $O_P(1/n^{3/2})$ term, equations (1) and (2) are the result.

Observe that the sum,

$$\Delta = \tfrac{1}{6}(1 - z^2)\frac{M_3}{V(\hat{p})} + \frac{z^2}{2} B , \tag{8}$$

in equation (6) has a simple estimator, $\delta_1 = (1/6 + z^2/3)b_1$. Strikingly, the two components of $\Delta$ have opposite signs when $z^2 > 0$. The first component captures the direct impact of the skewness of $\hat{p}$ on the Edgeworth expansion, but it is the second component, the result of replacing $V(\hat{p})$ by $v_1^*(\hat{p}) = v_1(\hat{p}) - B(\hat{p} - p)$ rather than $v_1(\hat{p})$, that dominates.

Unfortunately, it is not wholly legitimate to drop the $O_P(1/n^{3/2})$ term under the radical in equation (6) because it has the same asymptotic impact as other terms not dropped. (A discussion on keeping the $O(1/n^2)$ term $\Delta^2$ under the radical is deferred to Section 6.) Note that the source of the dropped $O_P(1/n^{3/2})$ term is the variance of $v_1^*(\hat{p})$. An *ad hoc* way of treating the instability of this idealized variance estimator is to estimate its *effective degrees of freedom*.

Consider the mean, $\overline{x}$, of *iid* normal variates, $x_1, \ldots, x_n$. The pivotal statistic $(\overline{x}-\mu)/\sqrt{v(\overline{x})}$, although asymptotically normal, has a Student *t* distribution with $n-1$ degrees of freedom. This degrees-of-freedom value is equal to 2 divided by the relative variance of the chi-squared random variable $v(\overline{x})$, the estimator for the variance of $\overline{x}$.

In an analogous fashion, we can replace standard normal distribution used to relate α to $z$ in equation (1) (through $z = \Phi^{-1}(\alpha)$) with a Student *t* distribution having effective degrees of freedom equal to 2 divided by the relative variance of $v_1{*}(\hat{p}) =$ $v_1(\hat{p}) - B(\hat{p}-p)$. The appendix shows this value to be

$$d = \frac{2\left[\sum_{h=1}^{H}\frac{W_h^2}{n_h}p_h(1-p_h)\right]^2}{\sum_{h=1}^{H}\frac{W_h^4}{n_h^3}\left[p_h(1-p_h)(1-2p_h)^2\frac{2n_h-3}{2(n_h-1)}+\frac{p_h(1-p_h)}{2(n_h-1)}\right]-\frac{\left[\sum_{h=1}^{H}\frac{W_h^3}{n_h^2}p_h(1-p_h)(1-2p_h)\right]^2}{\sum_{h=1}^{H}\frac{W_h^2}{n_h}p_h(1-p_h)}} \qquad (9)$$

or equivalently

$$d = \frac{2\left[\sum_{h=1}^{H}\frac{W_h^2}{n_h}p_h(1-p_h)\right]^2}{\sum_{h=1}^{H}\frac{W_h^4}{n_h^3}\left[p_h(1-p_h)(1-2p_h)^2+\frac{2\{p_h(1-p_h)\}^2}{(n_h-1)}\right]-\frac{\left[\sum_{h=1}^{H}\frac{W_h^3}{n_h^2}p_h(1-p_h)(1-2p_h)\right]^2}{\sum_{h=1}^{H}\frac{W_h^2}{n_h}p_h(1-p_h)}}. \qquad (10)$$

When all $n_h \geq 4$, $d$ in equation (9) can be consistently estimated by

$$\hat{d} = \frac{2\left[\sum_{h=1}^{H} \frac{W_h^2}{n_h - 1} \hat{p}_h (1-\hat{p}_h)\right]^2}{\sum_{h=1}^{H} W_h^4 \left[ g_h \frac{2n_h - 3}{2n_h^3 (n_h - 1)} + \frac{\hat{p}_h (1-\hat{p}_h)}{2n_h^2 (n_h - 1)^2} \right] - \frac{\left[\sum_{h=1}^{H} \frac{W_h^3}{(n_h - 1)(n_h - 2)} \hat{p}_h (1-\hat{p}_h)(1-2\hat{p}_h)\right]^2}{\sum_{h=1}^{H} \frac{W_h^2}{n_h - 1} \hat{p}_h (1-\hat{p}_h)}}, \quad (11)$$

where $g_h = \frac{n_h^3}{(n_h - 1)(n_h - 2)(n_h - 3)} \hat{p}_h (1-\hat{p}_h)(1-2\hat{p}_h)^2 - \frac{n_h}{(n_h - 1)(n_h - 3)} \hat{p}_h (1-\hat{p}_h)$

is an unbiased estimator for $p_h (1-p_h)(1-2p_h)^2$.

Equation (11) assumes the overall sample size is large, but allows the individual stratum sample sizes, the $n_h$, to be small. An alternative, slightly simpler, estimator for $d$ is consistent when all the $n_h$ are large, say at least 10 (formally, $1/n_h = O(1/n)$ for all $h$):

$$\hat{d}_A = \frac{2\left[\sum_{h=1}^{H} \frac{W_h^2}{n_h} \hat{p}_h (1-\hat{p}_h)\right]^2}{\sum_{h=1}^{H} \frac{W_h^4}{n_h^3} \hat{p}_h (1-\hat{p}_h)(1-2\hat{p}_h)^2 - \frac{\left[\sum_{h=1}^{H} \frac{W_h^3}{n_h^2} \hat{p}_h (1-\hat{p}_h)(1-2\hat{p}_h)\right]^2}{\sum_{h=1}^{H} \frac{W_h^2}{n_h} \hat{p}_h (1-\hat{p}_h)}}. \quad (12)$$

When the $p_h$ are roughly equal, an inconsistent, but more stable, estimator for $d$ in equation (10) is

$$\hat{d}_B = \frac{2\left[\sum_{h=1}^{H}\frac{W_h^2}{n_h}\right]^2 \hat{p}(1-\hat{p})}{\left\{\sum_{h=1}^{H}\frac{W_h^4}{n_h^3} - \frac{\left[\sum_{h=1}^{H}\frac{W_h^3}{n_h^2}\right]^2}{\sum_{h=1}^{H}\frac{W_h^2}{n_h}}\right\}(1-2\hat{p})^2 + \sum_{h=1}^{H}\frac{W_h^4}{n_h^3}\frac{2\hat{p}(1-\hat{p})}{(n_h-1)}}. \tag{13}$$

The last summation in the denominator of equation (13) can be viewed as a penalty for stratification. When all the $n_h$ are large, it can be ignored, and

$$\hat{d}_C = \frac{2\left[\sum_{h=1}^{H}\frac{W_h^2}{n_h}\right]^2 \hat{p}(1-\hat{p})}{\left\{\sum_{h=1}^{H}\frac{W_h^4}{n_h^3} - \frac{\left[\sum_{h=1}^{H}\frac{W_h^3}{n_h^2}\right]^2}{\sum_{h=1}^{H}\frac{W_h^2}{n_h}}\right\}(1-2\hat{p})^2}. \tag{14}$$

becomes a useful effective-degrees-of-freedom estimator.

Observe that when there is only one stratum, equation (12) collapses into equation (14) and the right-hand sides of both are infinite so long as $\hat{p}(1-\hat{p}) > 0$. When there are more than one strata, all the $n_h$ are large, and the $p_h$ are roughly equal, we can see from equation (14) that the effectively degrees of freedom are close to infinity for proportional allocation ($W_h = n_h/n$), and decrease as the variability of the stratum sampling fractions, the $n_h / N_h$, increases. The effective degrees of freedom also decrease and as $\hat{p}(1-\hat{p})$ get closer to 0.

Finally, when there is proportional allocation, the $p_h$ are roughly equal, and the $n_h$ are *not* all large, equation (11) collapses into

$$\hat{d}_B = \frac{n}{\sum_{h=1}^{H} \frac{n_h}{n} \frac{1}{(n_h - 1)}}.$$

If all $n_h \geq 3$, which is necessary to compute $\delta$ in equation (2.1), then $\hat{d}_B \geq 2n$.

Recalling that we need $n$ to be large (say $\geq$ 30) to use our coverage intervals in the first place, the observations above suggest that estimating the effective degrees of freedom will be most fruitful (that is to say, the estimated $d$ will be small enough to matter) when there is nonproportional allocation or widely divergent $p_h$, and $p$ is near either 0 or 1.

## 4. The *iid* model

One can avoid estimating the effective degrees of freedom entirely by assuming the *iid* model, as is natural in the single-stratum case. This is because the variance $\hat{p}$ can be written as $V(\hat{p}) = \left(\sum W_h^2 / n_h\right) p(1-p)$, which in turn can be estimated by $v_2(\hat{p}) =$

$\left(\sum W_h^2 / n_h\right) \hat{p}(1-\hat{p})$ instead of $v_1(\hat{p})$. An analogous alternative estimator for the third mean moment of $\hat{p}$ is $m_{3(2)} = \left(\sum W_h^3 / n_h^2\right) \hat{p}(1-\hat{p})(1-2\hat{p})$. In addition, since

$$\begin{aligned} \hat{p}(1-\hat{p}) - p(1-p) &= (\hat{p} - p) - (\hat{p}^2 - p^2) \\ &= (\hat{p} - p) - (\hat{p} - p)(\hat{p} + p) \\ &= (\hat{p} - p)(1 - \hat{p} - p) \\ &= (\hat{p} - p)(1 - 2p) - (\hat{p} - p)^2, \end{aligned}$$

and $\hat{p} - p = O_P(1/n^{1/2})$, $Cov\left(v_2(\hat{p}), \hat{p}\right) = \left(\sum W_h^2 / n_h\right) V(\hat{p})(1-2p) + O_P(n^{5/2})$.

As a result, $B = B_2 = Cov\left(v_2(\hat{p}), \hat{p}\right)/V(\hat{p})$ (which is no longer the same as $B = B_1 =$ $Cov\left(v_1(\hat{p}), \hat{p}\right)/V(\hat{p})$) can be estimated with $b_2 = \left(\sum W_h^2 / n_h\right)(1-2\hat{p})$.

When all the alternative estimators under the *iid* model given above are plugged into equation (6) with the unspecified $O_P(1/n^{3/2})$ terms dropped, equations (1) and (3) result. This version of the coverage intervals is better then the one using equation (2) − *assuming the iid model is correct* − because the $O_P(1/n^{3/2})$ term under the radical in equation (6) is, in fact, only $O_P(1/n^2)$. To see why this is so, observe that

$$\begin{aligned}
\hat{v}_2{*}(\hat{p}) - V(\hat{p}) &= v_2(\hat{p}) - b_2\left(\hat{p} - p\right) - V(\hat{p}) \\
&= \left(\sum W_h^2 / n_h\right)\hat{p}(1-\hat{p}) - \left(\sum W_h^2 / n_h\right)(1-2\hat{p})\left(\hat{p} - p\right) - \\
&\qquad \left(\sum W_h^2 / n_h\right) p(1-p) \\
&= \left(\sum W_h^2 / n_h\right)\left(\hat{p} - p\right)^2 \\
&= O_P(1/n^2).
\end{aligned}$$

Recall that Cai shows how to keep $O_P(1/n^2)$ terms in one-sided-interval construction for an unstratified *iid* population. The key is that $V(\hat{p})$ is expressible as a function of $p$. Although we could extend Cai's method for a stratified population under an *iid* model, we do not. We view the *iid* model as a convenient fiction useful when there are too few sampled units in a stratum or when $\hat{p}$ is very close to zero. Consequently, intervals based on the *iid* model are only rough approximations that we resist making too fine.

## 5. An Example

To see how our methods would work with a real stratified simple random sampling design, we divided a population of 6,000 units into three equal strata, and set the stratum sample sizes at $n_1 = 10$, $n_2 = 20$, and $n_3 = 30$. We let the overall proportion $p$ take on the values 0.001, 0.002, 0.003, ..., 0.998, 0.999 with $p_1 = p - p(1-p)$, $p_2 = p$, and $p_3 = p + p(1-p)$.

We generated a finite population of 2,000 unit values in each stratum $h$ and assigned values $x_{hi} = 1, 2, \ldots, 2{,}000$. We then drew 1,000 stratified random samples for each stratum sample size allocation. For each stratum proportion $p_h$, we let

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

The weighted estimate for the proportion of $y = 1$ was calculated for each value of $p$ and for each sample. Then simulated coverage probabilities were determined from the 1,000 samples for each $p$.

Figure 1 plots the coverages of one-sided 95% intervals computed using the general and *iid* models in equation (1) though (3) with $\Phi(.)$ in equation (1) denoting the *cdf* of a normal distribution for both methods. Since the "general" method is undefined when $\hat{p}$ is either 0 or 1, the "*iid*" method was used in its place in those circumstances.

Not surprisingly, given that the $p_h$ were not equal in this example, the general method provides coverages consistently closer to the nominal than the *iid*. The coverages using the general method are reasonably close to the nominal – between 92 and 98% − as long a $p$ is greater than 0.05 and less than 0.995 for the upper bound (i.e., the interval where

$p \leq \hat{p} + \delta + \sqrt{z^2 v(\hat{p}) + \delta^2}$ ) and greater than .005 and less than 0.90 for lower bound.

Observe that using either approach, when $p$ is very small (large) the coverage of the upper bound (lower-bound) is 100%. The reason for this is fairly obvious. From equation (1), it is clear the upper bound for $p$ is never less below 2δ, while the lower bound for $p$ is never less above 1 −2δ. Consequently, any $p$ value less than 2δ (greater than 1 −2δ) is in every upper-bound (lower-bound) coverage interval.

In fact, if all the $p_h$ were equal to $p$, and $p$ was less than or equal to 0.0485, or equivalently, $(1-p)^{60} \geq 0.05$, then $\hat{p}$ would have at least a 5% probability of being zero. Thus, no matter how small $\hat{p}$ was, any value of $p$ less than or equal to 0.0485 would have to be in the upper-bound interval to assure at least 95% coverage. Similarly, any value of $p$ greater than or equal to 0.9513 would have to be in every lower-bound interval to assure at least 95% coverage. As a consequence, finding an upper bound providing close to nominal 95% coverage when $p <$ 0.0485 or a lower bound when

$p >$ 0.9513 is a quixotic task. We have marked the beginnings of those regions in the plots with vertical "reference" lines.

We had hoped that changing Φ(.) in equation (1) into a Student's $t$ distribution with degrees of freedom set conservatively at the lesser of $\hat{d}_A$ and $\hat{d}_C$ in equations (12) and (14), respectively, would reduce the downward spikes in the coverages produced by the general method using the Normal distribution. As Figure 2 shows, mild downward spikes for some small $p$ values were removed from the plots of both the upper and lower bounds, but at the cost of over-coverages elsewhere. The deep downward spike in the lower-bound plot at 0.958 to the right reference line remained. At 0.951, the last evaluated $p$ value left of the reference line, the coverage using the $t$-distribution was only slightly closer to nominal than using the Normal distribution (86.8% vs. 86.1%). A second downward spike at 0.932, however, showed marked improvement (93.3% vs. 85.8%).

Comparisons of the methods proposed here with alternatives in the literature can be found in Liu and Kott (2007). The results there are generally encouraging, but, as here, the improvements realized from using a $t$-distribution rather than a Normal are modest.

## 6. Discussion

As noted in the introduction, one big difference between the coverage intervals proposed by Hall (1982) and the ones introduced in equation (1) is that the $\delta^2$ under the radical are missing from Hall's version. The reason for this is that the impact of the $\delta^2$ is asymptotically ignorable. They estimate $\Delta^2$ (see equation (8)), which is smaller than a term under the radical dropped in Section 3. Why, then, didn't we drop the $\delta^2$ as well? Theoretically, we didn't have to because they don't matter. Practically, their inclusion in equation (1) forces our coverage intervals to include 0 when $\hat{p} = 0$ and 1 when $\hat{p} = 1$, which Hall's does not. The empirical advantage of this is demonstrated in Liu and Kott (2007).

The problem with the asymptotics is that the positive floor put under the $p_h(1-p_h)$ to justify the Edgeworth expansion in equation (4) may not be reasonable. An alternative set of assumptions forcing $\tau$ to be o (1) but perhaps not as small as O($1/n$) could be developed to justify keeping the $\delta^2$. That task is left to the interested reader. One thing to note, however, is that when $1/\{n[p(1-p)]\}$ and thus $\tau$ are uncomfortably large, the fruit from using the Edgeworth expansion can be of dubious quality. That may be why the coverages of the upper (lower) bound for very small (large) $p$, as displayed in the graphs from the previous section and in Liu and Kott, are not as close to nominal as elsewhere.

It is a simple matter to construct a two-sided α-percent interval analogous to equation (1):

$$\hat{p} + \delta - \sqrt{z^2 v(\hat{p}) + \delta^2} \le p \le \hat{p} + \delta + \sqrt{z^2 v(\hat{p}) + \delta^2},$$

where $z = \Phi^{-1}(1/2 + \alpha/2)$ (note that $1/2 + \alpha/2 = 1 - (1-\alpha)/2$). When combined with equation (2) under the general model, the result is a coverage interval very similar to one proposed in Kott *et al.* (2001). The major difference is that $\delta$ is $(1/6 + z^2/3)b_1$ in equation (2.2), while it is effectively $(z^2/2)b_1$ in Kott *et al*.

Under simple random sampling, the two-sided interval proposed in Kott *et al*. is asymptotically identical to the Wilson interval. Brown *et al.* (2001) provides empirical support for the Wilson or score methodology, but Cai (2004) shows it is less well suited for constructing one-sided intervals. It over-covers when the Wald under-covers and under-covers when the Wald over-covers, both to a lesser degree.

Kott *et. al*. describes how to estimate $b_1$ – and thus for our purposes $\delta_1$ – under a stratified multi-stage design. Complications arise when the effective degrees of freedom for the idealized variance estimator, $v_1^*(\hat{p}) = v_1(\hat{p}) - \left[Cov\left(v(\hat{p}), \hat{p}\right)/V(\hat{p})\right]\left(\hat{p} - p\right)$, are not large and need to be estimated from the sample. Usually, the fewer the number of primary sampling units, the smaller the effective degrees of freedom. Estimating the effective degrees of freedom is especially difficult under a stratified multi-stage design because the population proportion can vary not only among the strata, but among and within the primary sampling units. Moreover, when there are adjustments for, say, nonresponse the probability that $y_i = 1$ can conceivably be related to the unit's weight.

To get a handle on these difficulties, one will very likely need to assume a working model for the $y_i$ values. The simplest is that they are independent and identically distributed, which, if true, would allow the use of the *iid* method developed in text (equations (1) and (3)) without further adjustment. Unfortunately, this will rarely be an effective strategy except when $p$ is very small or very large. Much empirical research is needed in this area.

Finally, it is a relatively simple matter to extend the one-sided-coverage-interval-methodology in equations (1) and (2) to a more general estimator than $\hat{p}$. Replacing $\hat{p}$ in equation (1) by a consistent estimator $\hat{t}$ for a population or model parameter $t$, we could likewise replace $v(\hat{p})$ in equation (2.1) by a consistent estimator for the variance of $\hat{t}$, $v(\hat{t})$, and

$$\frac{\sum_{h=1}^{H} \frac{W_h^3}{(n_h-1)(n_h-2)} \hat{p}_h(1-\hat{p}_h)(1-2\hat{p}_h)}{\sum_{h=1}^{H} \frac{W_h^2}{n_h-1} \hat{p}_h(1-\hat{p}_h)}$$

in equation (2.2) by a consistent estimator for both

$$\frac{E[(\hat{t}-t)^3]}{\left\{E[(\hat{t}-t)^2]\right\}} \quad \text{and} \quad \frac{Cov(v(\hat{t}), \hat{t})}{V(\hat{t})},$$

if such an estimator exists. Luckily, not only will both these terms be equal in many situations, they will often be very close to zero, so that the Wald methodology can be used. On the other hand, estimating the effective degrees of freedom of $v^*(\hat{t}) =$ $v(\hat{t}) - \left[Cov\left(v(\hat{t}), \hat{t}\right)/V(\hat{t})\right]\left(\hat{t}-t\right) \approx v(\hat{t})$ may remain a difficult exercise. See Kott (1994).

## Appendix: Estimating the Effective Degrees of Freedom for the Idealized Variance Estimator

Suppose, as in the text, a set of independent Bernoulli random variables $y_i$ with a common mean within each of $H$ strata. Let $p_h$ be the (super)population mean of the $y_i$ in stratum $h$, and $\hat{p}_h$ is the sample mean within stratum $h$. Now $\hat{p} = \sum W_h \hat{p}_h$ is an unbiased estimator for $p = \sum W_h p_h$ where $W_h = N_h/N$, and $N_h$ is the finite population size of stratum $h$. In what follows, we assume $n_h \geq 4$ in every stratum.

The variance of $\hat{p}$ under the model described above, $V = \sum W_h^2 p_h(1-p_h)/n_h$ (denoted $V(\hat{p})$ in the text), can be estimated in an unbiased fashion by $v = \sum W_h^2 \hat{p}_h(1-\hat{p}_h)/(n_h-1)$ (denoted $v_1(\hat{p})$ in the text). Moreover,

$$E[(\hat{p}-p)v] = Cov(\hat{p}, v) = \sum W_h^3 p_h(1-p_h)(1-2p_h)/n_h^2 \qquad \text{(A.1)}$$

is likewise unbiasedly estimated by

$$e[(\hat{p}-p)v] = \sum W_h^3 \hat{p}_h(1-\hat{p}_h)(1-2\hat{p}_h) / [(n_h-1)(n_h-2)]. \qquad \text{(A.2)}$$

Now some work shows that the variance of $v$ is

$$Var(v) = \sum W_h^4 E\{ [\hat{p}_h(1-\hat{p}_h)/(n_h-1) - p_h(1-p_h)/n_h]^2\}$$

$$= \sum W_h^4 E\{ [(p_h + d_h)(1-p_h-d_h)/(n_h-1) - p_h(1-p_h)/n_h]^2\}$$

$$= \sum W_h^4 E\{\ [(1-2p_h)d_h/(n_h-1) - d_h^2/(n_h-1) + p_h(1-p_h)/(n_h[n_h-1])\ ]^2\}$$

$$= \sum W_h^4\{\ (1-2p_h)^2 p_h(1-p_h)/[n_h(n_h-1)^2] - 2(1-2p_h)^2 p_h(1-p_h)/[n_h^2(n_h-1)^2]$$

$$+ \{(1-2p_h)^2 p_h(1-p_h)/[n_h^3(n_h-1)^2] + 2[p_h(1-p_h)]^2/[n_h^3(n_h-1)]\}$$

$$= \sum W_h^4\{\ (1-2p_h)^2 p_h(1-p_h)/n_h^3 + 2[p_h(1-p_h)]^2/[n_h^3(n_h-1)]\} \qquad \text{(A.3)}$$

where $d_h = \hat{p}_h - p_h$, $E(d_h^2) = p_h(1-p_h)/n_h$ , $E(d_h^3) = (1-2p_h)\ p_h(1-p_h)/n_h^2$,

$E(d_h^4) = p_h(1-p_h)(\ 1-3p_h + 3p_h^2)/n_h^3 + 3[p_h(1-p_h)]^2\ (n_h-1)/n_h^3$, and so

$E(d_h^4) - [p_h(1-p_h)/n_h]^2 = (1-2p_h)^2 p_h(1-p_h)/n_h^3 + 2\ [p_h(1-p_h)]^2\ (n_h-1)/n_h^3$.

These come from the independence of the $y_i - p_h$ across all $i$ and the moments:

$\mathrm{E}(y_i - p_h) = 0$, $\mathrm{E}(y_i - p_h)^2 = p_h(1-p_h)$, $\mathrm{E}(y_i - p_h)^3 = p_h(1-p_h)^2 + (1-p_h)\ p_h^2 =$

$p_h(1-p_h)\ (1-2p_h)$, $\mathrm{E}(y_i - p_h)^4 = p_h(1-p_h)^3 + (1-p_h)\ p_h^3 = p_h(1-p_h)(\ 1-3p_h + 3p_h^2)$.

for $i$ in stratum $h$.

For estimation purposes, it is helpful to rewrite $Var(v)$ as

$$Var(v) = \sum W_h^4[p_h(1-p_h)/n_h^3]\ \{(1-2p_h)^2 + 2p_h(1-p_h)/(n_h-1)\}$$

$$= \sum W_h^4[p_h(1-p_h)/n_h^3]\ \{(1-2p_h)^2 + [1-(1-2p_h)^2]/[2(n_h-1)]\}$$

$$= \sum W_h^4[p_h(1-p_h)/n_h^3]\ \{(1-2p_h)^2(2n_h-3)/(2n_h-2)] + 1/(2n_h-2)\}$$

$$= \sum (W_h^4/n_h^3)\{\ [p_h(1-p_h)(\ 1-2p_h)^2(2n_h-3)/(2\mathrm{n}_h-2)]] + p_h(1-p_h)/(2n_h-2)\}$$

$$= \sum (W_h^4/\mathrm{n_h}^3)\{\ G_h(2n_h-3)/(2n_h-2) + p_h(1-p_h)/(2n_h-2)\}, \qquad \text{(A.4)}$$

where $G_h = p_h(1-p_h)(\ 1-2p_\mathrm{h})^2$.

An estimator for $p_h(1-p_h)$ is $n_h\,\hat{p}_h(1-\hat{p}_h)/(n_h-1)$. An unbiased estimator for $G_h$ is $g_h =$

$\{n_h^3\,\hat{p}_h(1-\hat{p}_h)(1-2\,\hat{p}_h)^2/[(n_h-1)(n_h-2)(n_h-3)]\} - n_h\,\hat{p}_h(1-\hat{p}_h)/[(n_h-1)(n_h-3)]$.

To see why, first note that $G_h = -4p_h^4 + 8p_h^3 - 5p_h^2 + p_h$, and

$g_h = n_h^3\{(-4\,\hat{p}_h^{\,4} + 8\,\hat{p}_h^{\,3} - 5\,\hat{p}_h^{\,2} + \hat{p}_h)/[(n_h-1)(n_h-2)(n_h-3)]\} - n_h(\hat{p}_h - \hat{p}_h^{\,2})/[(n_h-1)(n_h-3)]$

Observe

$$
\begin{aligned}
E(\hat{p}_h^{\,4}) &= E\{(p_h + d_h)^4\} \\
&= p_h^4 + 4p_h^3\mathrm{E}(d_h) + 6p_h^2\mathrm{E}(d_h^2) + 4p_h\mathrm{E}(d_h^3) + \mathrm{E}(d_h^4) \\
&= p_h^4 + 6p_h^3(1-p_h)/n_h + 4p_h^2(1-p_h)(1-2p_h)/n_h^2 + \\
&\qquad p_h(1-p_h)(1-3p_h+3p_h^2)/n_h^3 + 3(n_h-1)[p_h(1-p_h)]^2/\mathrm{n_h}^3 \\
&= p_h^4\{1-6/n_h + 8/n_h^2 - 3/n_h^3 + 3/n_h^2 - 3/n_h^3\} + \\
&\qquad p_h^3\{6/n_h - 12/n_h^2 + 6/n_h^3 - 6/n_h^2 + 6/n_h^3\} + \\
&\qquad p_h^2\{4/n_h^2 - 4/n_h^3 + 3/n_h^2 - 3/n_h^3\} + p_h/n_h^3 \\
&= n_h^{-3}\{p_h^4(n_h-1)(n_h-2)(n_h-3) + 6p_h^3(n_h-1)(n_h-2) + 7p_h^2(n_h-1) + p_h\},
\end{aligned}
$$

$$
\begin{aligned}
E(\hat{p}_h^{\,3}) &= \mathrm{E}\{(p_h + d_h)^3\} \\
&= p_h^3 + 3p_h^2(1-p_h)/n_h + p_h(1-p_h)(1-2p_h)/n_h^2 \\
&= p_h^3\{1-3/n_h + 2/n_h^2\} + p_h^2\{3/n_h - 3/n_h^2\} + p_h/n_h^2 \\
&= n_h^{-2}\{p_h^3(n_h-1)(n_h-2) + 3p_h^2(n_h-1) + p_h\},
\end{aligned}
$$

$$
\begin{aligned}
E(\hat{p}_h^{\,2}) &= \mathrm{E}\{(p_h + d_h)^2\} \\
&= p_h^2 + p_h(1-p_h)/n_h \\
&= n_h^{-1}\{p_h^2(n_h-1) + p_h\},
\end{aligned}
$$

and $\quad E(\hat{p}_h) = p_h$.

The proof that $E(g_h) = G_h$ is now straightforward.

An unbiased estimator for $Var(v)$ in equation (A.4) is then

$$var(v) = \sum W_h^4\{ g_h(2n_h - 3)/[n_h^3(2n_h - 2)] + \hat{p}_h(1 - \hat{p}_h)/[n_h^2(n_h - 1)(2n_h - 2)]\}. \quad \text{(A.5)}$$

Let $v^* = v - (\hat{p} - p)E[v(\hat{p} - p)]/V$ be the idealized variance estimator for $\hat{p}$ proposed by Anderson and Nerman. It effective degrees of freedom (2 divided by its relative variance) is

$$d = \frac{2v^2}{Var(v) - \frac{\{E[v,(\hat{p} - p)]\}^2}{v}},$$

where equations (A.1) and (A.4) (or (A.1) and (A.3)) provide values for $E[v,(\hat{p} - p)]$ and $Var(v)$ respectively.

A consistent estimator for $d$ under mild conditions we assume to hold is

$$\hat{d} = \frac{2v^2}{var(v) - \frac{\{e[v,(\hat{p} - p)]\}^2}{v}},$$

where equations (A.2) and (A.5) provide values for $e[v,(\hat{p} - p)]$ and $var(v)$ respectively.

## References

Andersson, P.G. and Nerman, O. (2000). A balanced adjusted confidence interval procedure applied to finite population sampling. Presented at the International Conference of Establishment Surveys, Buffalo, NY.

Brown, L.D., Cai, T. and Dasgupta, A. (2001). Interval estimation for a binomial poportion. *Statistical Science*, 16, 101-133.

Cai, T. (2004). One-sided confidence intervals in discrete distributions. *Journal of Satistical Planning and Inference*, 131, 63-88.

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

Kott, P. S. (1994). Hypothesis testing of linear regression coefficients with survey data. Survey Methodology, 20, 159-164.

Kott, P. S., Anderson, P. G. and Nerman O. (2001). Two-sided coverage intervals for small proportion based on survey data. Presented at Federal Committee on Statistical Methodology Research Conference, Washington, DC.

Liu, Y. K. and Kott, P. S. (2007). Evaluating alternative one-sided coverage intervals for a binomial proportion. Submitted to the *Journal of Official Statistics*.

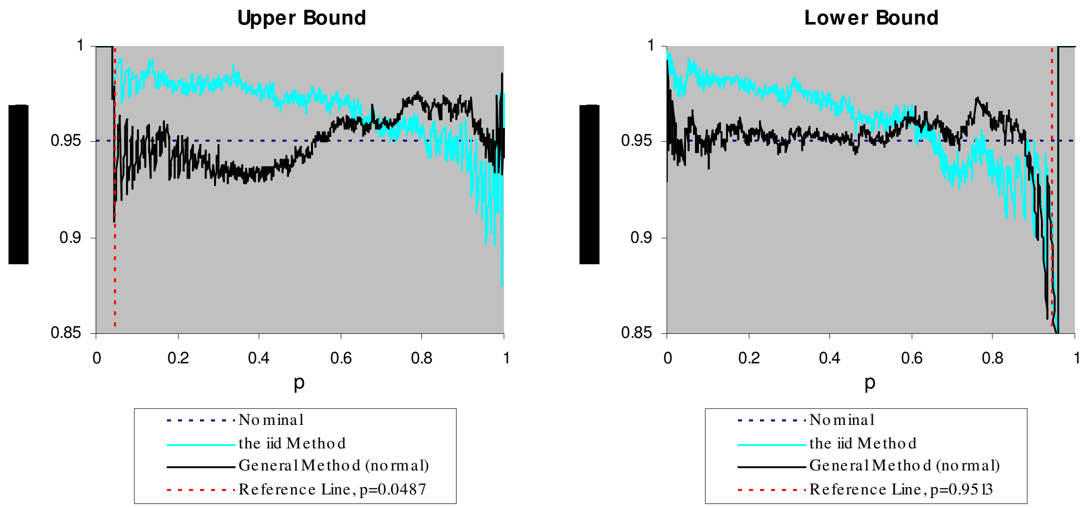


**Figure 1.** Simulated Coverages for One-Sided 95% Bounds:
Methods Based on the General and *iid* Models
Using the Normal Distribution

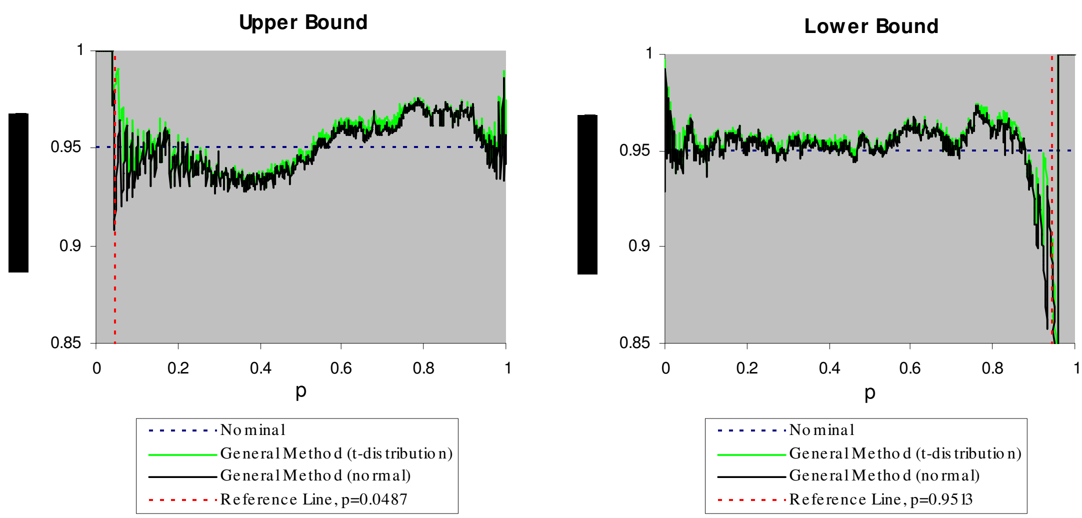


**Figure 2.** Simulated Coverages for One-Sided 95% Bounds:
Methods Based on the General Model
Using the Normal and Student *t* Distributions

**Evaluating Alternative One-Sided Coverage Intervals for a Binomial Proportion**

Yan K. Liu[1] and Phillip S. Kott[2]

The construction of coverage intervals for a binomial proportion is difficult, especially when the proportion is very small or very large. Most of the methods treated in the literature implicitly assume simple random sampling. These interval-construction 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 binomial-proportion coverage intervals 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. Statistics of Income Division, IRS, P.O.Box 2608, Washington, DC 20013, USA. Email:yan.k.liu@irs.gov

2. Research and Development Division, National Agricultural Statistics Service, 3251 Old Lee Highway, Fairfax, VA 20250, USA. Email: pkott@nass.usda.gov.

The authors express a special note of thanks to Kevin Cecco at IRS for his encouragement and initial direction at the beginning of this project.

## 1. Introduction

Because of the poor performance of the standard Wald method for constructing coverage (confidence) intervals around 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 an even more difficult task. Cai (2004) and Hall (1981) use an Edgeworth expansion to develop one-sided coverage intervals under a simple a random sample. Kott and Liu (2007) modifies the Hall method and extends it to handle data from a complex sample design with a particular emphasis on stratified (simple) random sampling.

We are particularly 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 summary and discussion or our results.

## 2. Interval Construction Under a Simple Random Sample

Let X follow a binomial distribution with parameters $n$ and $p$. The parameter $p$ is 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 one-sided $(1-\alpha)\%$ coverage intervals. For 95% coverage intervals, $\alpha = 0.05$, and 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): $(\hat{p} - p)\big/\sqrt{v(\hat{p})} \rightarrow N(0,1)$, where $v(\hat{p}) = \hat{p}(1-\hat{p})/(n-1)$. The lower and upper bounds are

$$L_S = \hat{p} - z\sqrt{\hat{p}(1-\hat{p})/(n-1)}, \quad \text{and}$$
$$U_S = \hat{p} + z\sqrt{\hat{p}(1-\hat{p})/(n-1)}. \tag{1}$$

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

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

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

where

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

*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

$$L_L = \frac{e^{\lambda_L}}{1+e^{\lambda_L}}, \quad \text{where } \lambda_L = \hat{\lambda} - z\sqrt{v(\hat{\lambda})}\text{, and}$$

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

*Arcsine(root) Interval*

Another transformation-stabilizing variance is the arcsine(root) transformation, $\delta = \arcsin(\sqrt{p})$.

The interval for $\delta$ 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 $p$:

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

*Jeffrey Interval*

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

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

*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:

$$L_{CP} = \left\{ p : \sum_{t=0}^{x-1} \binom{n}{t} p^t (1-p)^{n-t} = 1 - \alpha \right\}, \text{ and}$$

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

They can be expressed in terms of Beta distribution as

$$L_{CP} = \text{Beta}(\alpha; x, n-x+1)\text{, and}$$
$$U_{CP} = \text{Beta}(\alpha; x+1, n-x)\,. \tag{6}$$

<u>*Mid-P Clopper-Pearson Interval*</u>

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

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

The interval can be expressed in terms of Beta distribution as

$$L_{MP} = \frac{1}{2}\left\{\text{Beta}\left(\alpha; x, n-x+1\right) + \text{Beta}\left(\alpha; x+1, n-x\right)\right\}\text{, and}$$
$$U_{MP} = \frac{1}{2}\left\{\text{Beta}\left(1-\alpha; x, n-x+1\right) + \text{Beta}\left(1-\alpha; x+1, n-x\right)\right\}. \tag{7}$$

Note its similarity to the Jeffrey interval in equation (5).

*Brown et al.* (2001) evaluates the properties of these seven methods for constructing two-sided intervals (replacing α by α/2 and z by the z-score of 1−α/2). 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

$$L_H = \hat{p} + \delta - z\sqrt{v(\hat{p})}\text{, and}$$
$$U_H = \hat{p} + \delta + z\sqrt{v(\hat{p})}\text{,} \tag{8}$$

where

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

The translation term, $\delta$, 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) goes further than Hall in correcting for the skewness in $\hat{p}$ by keeping $O_P(1/n^2)$ terms producing the bounds:

$$L_{Cai} = \breve{p} - \frac{z}{\sqrt{n}}\sqrt{\hat{p}(1-\hat{p}) + \frac{\gamma_1 \hat{p}(1-\hat{p}) + \gamma_2}{n}}\text{, and}$$
$$U_{Cai} = \breve{p} + \frac{z}{\sqrt{n}}\sqrt{\hat{p}(1-\hat{p}) + \frac{\gamma_1 \hat{p}(1-\hat{p}) + \gamma_2}{n}}\text{,} \tag{9}$$

where

$$\breve{p} = \frac{\hat{p} + \eta / n}{1 + 2\eta / n},\ \eta = \frac{z^2}{3} + \frac{1}{6},\ \gamma_1 = -\frac{13}{18} z^2 - \frac{17}{18} \text{ and } \gamma_2 = \frac{1}{18} z^2 + \frac{7}{36}.$$

*Kott-Liu Interval*

Under simple random sampling, Kott and Liu (2007) proposes a slight modification of the Hall interval that better handles samples with small $\hat{p}(1-\hat{p})$ values:

$$L_{KL} = \hat{p} + \delta - \sqrt{z^2 v(\hat{p}) + \delta^2}\text{, and}$$
$$U_{KL} = \hat{p} + \delta + \sqrt{z^2 v(\hat{p}) + \delta^2}, \tag{10}$$

where $v(\hat{p})$ and $\delta$ are unchanged. Notice that the lower bound attains its minimum value, 0, when $\hat{p} = 0$, and the upper bound attains its maximum value, 1, when $\hat{p} = 1$. This method will be described further in the following section.

*Other Intervals*

There are also various continuity-correction approaches (Vollset 1993) that are not included in this paper. Two other methods not treated here are the Wilson-logit and likelihood-ratio interval (Feng 2006). These methods employ an iteration algorithm to obtain the interval end-points. Finally, 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 $\lambda = np$ (Newcombe 1998, Feng 2006). The lower and upper bounds for $p$ are

$L_P = \chi^2_{2x,\alpha} \big/ (2n)$, and

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

This method has to be redefined for $p$ near 1 to be effective and is not useful when $p$ is not very near either 0 or 1.

*2.2 Comparison of One-Sided Intervals Under Simple Random Sampling*

In this subsection, the methods defined in equations (1) through (10) 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 $D(x) = \begin{cases} |L(x) - p|, \text{ for the lower bound} \\ \\ |U(x) - p|, \text{ for the upper bound} \end{cases}.$

We are interested in a setting where the sample size $n$ is relatively small but large enough for the asymptotic theory supporting some of the methods to be effective. Therefore, we evaluate a sample of size 30. Coverages perform differently for different sample sizes and different values of $p$ (Brown *et al.*, 2001 and 2002, discuss this at length for two-sided intervals). Thus, we evaluate one-sided coverages over the entire range of potential $p$ values.

We make a few sensible modifications of the methods when $x = 0$ or 1. We force the lower bound to be 0 at $x = 0$ and the upper bound to be 1 at $x = 1$. In addition, when a bound is not defined at $x = 0$ or 1 for a method (the Wald, Logit and Mid-P), we take a conservative tact and replace it with the Clopper-Pearson.

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, say, $p < 0.2$ 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, …, 0.998, 0.999.

The plots of coverage probabilities are displayed in Figure 1. The vertical lines at $p = 0.905$ represents the $p$-value where $p^{30} = 0.05$. This mean when $p \geq 0.905$, $\hat{p}$ has at least a 5% probability of being 1.

The following conclusions can be drawn from the plots in Figure 1:

- The Wald and Arcsine methods are systematically biased, sometimes in one direction sometimes in the other.
- The Clopper-Pearson method always has at least the nominal coverage (95%), but often over-covers. It has 100% coverage when $p \geq 0.905$.
- The Wilson and Logit methods are systematically biased in the opposite direction of the Wald but to a lesser degree. They tend to under-cover for small $p$ and over-cover for large $p$. The over-coverage for the Wilson near $p = 1$ is not as pronounced as for the Clopper-Pearson.
- The Jeffrey and Hall methods have large downward spikes (under-coverages) near the two boundaries.
- The Mid-P has some large downward spikes near $p = 0$, but performs reasonably well for large $p$.
- The Kott-Liu and Cai methods provide good coverages almost everywhere. Both have 100% coverages as $p$ gets very close to 1, but this "lip" begins for the Kott-Liu (at 0.929) while the Cai is still experiencing its worst downward spike or "dip" (it reaches a minimum coverage of 87.1% before beginning its lip at 0.935; the Kott-Liu minimum coverage is 89.4%). Before then, the methods have identical coverages for large $p$ values ($\geq .85$).

Analogous graphs for $n = 20$, 30, 60, and 120 (not show) behave similarly.

We plot the average distances of lower bounds versus the tail values of $p$ for the better methods (Mid-P, Cai, and Kott-Liu) and for the conservative Clopper-Pearson in Figure 2. In general, the average distance is longer when the coverage probability is larger. The Clopper-Pearson has a much longer average distance than the other methods, not surprising since it tends to be conservative. For small $p$, the Kott-Liu and Cai behave very similarly. For large $p$, the Kott-Liu tends to be slightly longer than the Cai. The Mid-P becomes longer than Kott-Liu and Cai when $p$ gets near 1 but not before.

In summary, the Kott-Liu and Cai methods are the best in terms of having coverages almost always close to the nominal. The Clopper-Pearson, never under-covers, but has longer average distances. Many view the property of never providing less than nominal coverage as very desirable, if not absolutely required (see the discussions in Brown *et. al*, 2001). They argue that the user should have "confidence" that his/her interval always covers at least as well as advertised. Such confidence is rarely justified with complex-sample data, as we shall see.

## 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 $\hat{p} = \sum_s w_k x_k / \sum_s w_k$ .

### *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 (estimated) *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

$$n^* = \frac{n}{DEFF(\hat{p})} = \frac{\hat{p}(1-\hat{p})}{v(\hat{p})} \tag{11}$$

Sometimes, $n^*$ is defined as 1 plus the left-hand side of equation (11). The distinction is usually trivial when $n \geq 30$.

The *idealized effective sample size* $\tilde{n}$ features the true variance $V(\hat{p})$ in the denominator of equation (11) in place of the estimated variance $v(\hat{p})$. Unfortunately, $V(\hat{p})$ is unknown and needs to be estimated from the sample in practice.

The *ad hoc* procedure of replacing $n$ by $n^*$ is 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) treats a few other intervals with this procedure.

We focus in this section on an empirical evaluation of one-sided interval 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 actually discuss three different coverage intervals for data from a stratified random sample.

*Basic Kott-Liu Interval*

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

$$U_{KL1} = \hat{p} + \delta_1 + \sqrt{z^2 v_1(\hat{p}) + \delta_1{}^2} \text{ ,} \qquad (12)$$

where $v_1(\hat{p}) = \sum_{h=1}^{H} W_h^2 \hat{p}_h (1-\hat{p}_h) \big/ (n_h - 1)$,

and

$$\delta_1 = \left( \frac{z^2}{3} + \frac{1}{6} \right) \frac{\sum^H W_h^3 \hat{p}_h (1-\hat{p}_h)(1-2\hat{p}_h) \big/ \big[(n_h-1)(n_h-2)\big]}{\sum^H W_h^2 \hat{p}_h (1-\hat{p}_h) \big/ (n_h-1)} \ . \qquad (13)$$

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. The estimation has its own random error, which cannot be completely eliminated from the Edgeworth expansion (moreover, following Cai and keeping $O_P(1/n^2)$ terms becomes impossible).

*DF-adjusted Kott-Liu Interval*

One way to adjust for the error in the implicit estimator for $V(\hat{p})$ in the basic Kott-Liu method is by that replacing the $z$-score in equation (12) with a $t$-score from a Student $t$. A $t$-distribution needs a degrees-of-freedom calculation. Kott and Liu discusses 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 $a_1 = \sum^H W_h^2 \hat{p}_h (1-\hat{p}_h) \big/ n_h$ , $a_2 = \sum^H W_h^3 \hat{p}_h (1-\hat{p}_h)(1-2\hat{p}_h) \big/ n_h^2$ , and $a_3 = \sum^H W_h^4 \hat{p}_h (1-\hat{p}_h)(1-2\hat{p}_h)^2 \big/ n_h^3$ .

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

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

A slightly conservative policy, followed here, 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 (12).

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

If an independent and identically distributed (*iid)* Bernoulli model is assumed, then a different way to generalize equation (10) is with

$$L_{KL2} = \hat{p} + \delta_2 - \sqrt{z^2 v_2(\hat{p}) + \delta_2{}^2} \text{, and}$$
$$U_{KL2} = \hat{p} + \delta_2 + \sqrt{z^2 v_2(\hat{p}) + \delta_2{}^2} \text{,} \tag{14}$$

where $v_2(\hat{p}) = \sum^H W_h^2 \hat{p}(1-\hat{p}) \big/ n_h$ , and

$$\delta_2 = \left( \frac{1-z^2}{6} \frac{\sum^H W_h^3 / n_h^2}{\sum^H W_h^2 / n_h} + \frac{z^2}{2} \sum^H \frac{W_h^2}{n_h} \right)(1-2\hat{p}) \,. \tag{15}$$

Since both the basic and DF-adusted Kott-Liu intervals are undefined when $\hat{p}$ = 0 or 1, Kott and Liu suggests using the *iid* method in equation (14) in this situation.

*3.2 Comparison of One-Sided Intervals under Stratified Random Sampling*

All the methods described in the text are evaluated under the following stratified random sampling designs 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. We consider these six settings for the stratum sample sizes and the comparative values of the $p_h$. They are shown in Table 1.

**Table 1. Simulation Settings**

| Stratum Sample Sizes $n_1, n_2, n_3$ | Stratum Binomial Proportions $p_1, p_2, p_3$ | |
|---|---|---|
| | $p, p, p$ | $p-p(1-p), p, p+p(1-p)$ |
| 10, 10, 10 | A | B |
| 10, 30, 10 | C | D |
| 10, 10, 30 | [same as C] | E |
| 30, 10, 10 | [same as C] | F |

One sample size allocation – 10, 10, 10 – has a total sample size of 30, our minimum. The other allows one strata to be big enough to stand alone, $n_h$ = 30, while the other two strata contain 10 samples. As for the $p_h$ values, they are either all equal or their spread is, in some sense, maximized while allowing all the $p_h$ values to fall into the 0 to 1 range.

For 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 (10, 10, 30 mirrors 30, 10, 10). Due to the space limitation, we only display the lower-bound coverage plots using the Mid-P, Clopper-Pearson, Cai, and three Kott-Liu methods.

The plots for setting A (not displayed) mirror those in Figure 1 with the three Kott-Liu methods being virtually identical. This is not surprising since the $p_h$ are equal, the idealized effective samples size is 30, and the effective degrees of freedom are nearly infinite (as in simple random sampling).

Figure 3 displays the plots for Setting B. Despite the variability in the $p_h$, not much changes from Setting A. The Clopper-Pearson has a small dip below 95% (to 94.7%), but that occurs with Setting A as well (not shown), probably due to the effective sample size not being estimated exactly. Its lip again begins at 0.905, which is marked in all the plots.

The basic and DF-adjusted Kott-Liu methods remain virtually identical everywhere, while the *iid* version is slightly more variable than the others when $p$ is roughly between 0.2 and 0.8 but matches their behavior in the tails. The Mid-P is similar to the three Kott-Liu methods when $p > 0.8$ for both settings A and B, but continues to be plagued by downward spikes for some very small $p$.

Since the Kott-Liu *iid* method may have problems when the $p_h$ are not equal, Figures 4, 5, and 6 display the coverage plots for Settings D, E, and F. For Setting D, the basic Kott-Liu has a very final deep dip just before its lip. The DF-adjusted version is only slightly better. Its lip starts at 0.952 rather than at 0.956 (the basic has a minimum coverage of 79.4%, the DF-adjusted 81.4%). The Kott-Liu *iid* method hardly dips at all. Its lip starts at 0.948.

The lip for the Mid-P starts at 0.942, just like the Clopper-Pearson. This corresponds to the $p$-value such that $p^{50} = 0.05$, which is marked by a vertical line in all the plots. The Cai's lip doesn't begin until at 0.961, while its dip (bottoming at 87.5%) is not as great as the basic and DF-adjusted Kott-Liu methods.

In Setting E, all the methods suffer from a deep dip before the final lip. Here, there is no advantage of the DF-adjusted Kott-Liu over the basic. Its lip starts slightly earlier, but by then the basic's dip has ended. The Clopper-Pearson has the slightest dip and longest lip among the methods, but its dip is well below the nominal (87.8% at 0.941 as opposed to *iid* Kott-Liu's 84.4% at 0.947). The Cai has the deepest dip (74.1% as opposed to the basic and DF-adjusted

Kott-Liu's 83.1%). Both the Clopper-Pearson and the Kott-Liu *iid* method consistently over-cover when $p$ is less than 0.5.

In Setting F, only the basic and DF-adjusted Kott-Liu methods have final dips, and these are modest (the basic's bottom is 88.8% as 0.955, while the DF-adjusted in 91.1% at 0.951). The Clopper-Pearson consistently over-covers for all values of $p$. The Kott-Liu *iid* method consistently over-covers when $p$ is greater than 0.5 and suffers downward spikes for very low values of $p$, but not as severe as the Mid-P. The Mid-P and Cai tend to over-cover for $p$ >0.6, but by not as much as the Clopper-Pearson and Kott-Liu *iid* methods.

The average distances for tail $p$-values in Settings B and C are displayed in Figures 7 and 8, respectively, for the Clopper-Pearson, Cai, DF-adjusted Kott-Liu and Kott-Liu i*id* methods. The conservative Clopper-Pearson method exhibits the longest average distances, while the Cai method tends to have the smallest average distances, but not by much. The average-distance plots for the other settings (not displayed) are similar.

## 4. Summary and 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 reviewed, the Cai and Kott-Liu produced one-sided intervals coverages closest to nominal. We also confirmed that the Clopper-Pearson method always provided at least the nominal coverage, which many find a singularly desirable property.

We then turned to stratified random sampling. We adjusted all the non-Kott-Liu methods by replacing the sample size with an estimate for the effective sampling size. The Clopper-

Pearson was still the most conservative method with coverage probabilities usually, *but not always*, at or above the nominal level. The potential for under-coverage was larger when the sampling fraction varied across the strata.

The basic Kott-Liu method worked reasonably well for constructing lower bounds when $p < 0.9$ (and symmetrically upper bounds when $p > 0.1$), but lower bounds often under-covered for larger $p$. This under-coverage was less severe when the sampling fractions were equal across the strata.

The Cai method appeared to have strengths and weaknesses in coverage similar to the basic Kott-Liu method and often slightly smaller average lengths. Forcing the lower bound to be zero when $\hat{p}$ was zero removed what would have been sharp downward spikes for small $p$ values.

Adjusting the basic Kott-Liu method for its effective degrees of freedom sometimes improved coverages for large and small $p$, but not by much. Based on our empirical analysis, it appears that the *iid* version of the Kott-Liu provides a better, if not perfect, alternative for lower-bound (upper-bound) construction when $p$ is large (small). Another alternative is suggested below.

The lower bounds constructed using any of the methods have "lips" very near 1. That is, a region in which coverage is 100%. It is easy to see that this region includes all $p > 1-2\delta_1$ (see equations (12) and (13)) using the basic Kott-Liu method and all $p > 1-2\delta_2$ using the *iid* method (see equations (14) and 15)).

Using the Clopper-Pearson method, the lip begins in general at $p = p_L$, where $p_L{}^n = \alpha$, or equivalently,

$$p_L = p_L(\alpha, n) = \exp[\log(\alpha)/n]. \quad (16)$$

Suppose all the $p_h$ were equal to, say, $r$. If $r$ were greater or equal to $p_L$, and thus in the Clopper-Pearson lip, then $\hat{p}$ would have at least a probability α of being 1. No matter how large $\hat{p}$ was, $r$ would have to be in the lower one-sided interval to assure at least (1 −α)% coverage. As a consequence, finding a lower bound producing close to the nominal (1−α)% coverage when $p = r$ can be an impossible task. Nevertheless, it would be a prudent rule not to let the lower bound for an interval be any higher than $p_L$ (and, symmetrically, not let the upper bound be any lower than $p_U = 1 - p_L$ ). The size of the lip from using this rule is of asymptotic order $1/n$: it decreases as the sample size increases.

We have marked where $p_L$ falls in our coverage plots. Notice that not allowing the lower bound to be higher than $p_L$ reduces the size of dips that would result from using the Cai or one of the Kott-Liu methods in the settings displayed in Figures (1) and (3). There remain deep dips using all the methods in Setting E (Figure 5), even the Clopper-Pearson. This may be because the $p_h$ are not all equal and neither are the sampling fractions.

Observe that when the sampling fractions are the same across the three strata:

$$\log(p_1^{n_1} p_2^{n_2} p_3^{n_3}) = \sum n_h \log\{p[1+(p_h - p)/p)]\} \approx \sum n_h \{\log(p) + (p_h - p)/p\} = \log(p) = \log(p^n),$$

so the impact of the variability of the $p_h$ is muted. This suggests the following policy when the sampling fractions are not all equal: setting the maximum value of the lower bound at $p_{L2} = p_L(\alpha, N \min\{n_h/N_h\})$ with $p_L(.,.)$ defined in equation (16) (and setting the minimum value of the upper bound at $1 - p_L(\alpha, N \min\{n_h/N_h\})$. Such a policy will often be very conservative, extending the region where coverage will be 100%. This is a reflection of the difficulty of constructing a lower bound at all when $p > p_L(\alpha, N \min\{n_h/N_h\})$, and the variability among the

$p_h$ is unknown. *There is no parallel difficulty constructing an upper bound for large p or a lower bound for small p.* In any event, when one-sided coverage intervals for a small or a large proportion is a survey goal, it would be wise to avoid stratification schemes with widely varying sampling fractions if possible.

Constructing one-sided coverage intervals from samples derived using a stratified, multi-stage design were not addressed in this paper, but the Kott-Liu methods (perhaps modified in the tails) can in principle, be extended to cover such samples. See Kott *et al.* (2001) for a method of estimating the replacement for $\sum^H W_h^3 \hat{p}_h(1-\hat{p}_h)(1-2\hat{p}_h) \big/ \left[(n_h-1)(n_h-2)\right]$ in equation (13) under a stratified, multi-stage sample; $\sum^H W_h^2 \hat{p}_h(1-\hat{p}_h) \big/ (n_h-1)$ can be replaced by a standard randomization-based variance estimator for $\hat{p}$. More work on data from such designs will have to wait for another time.

## 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, 30, 160-201.

Cai, T. (2004). One-sided Confidence Intervals in Discrete Dstributions. Journal of Statistical Planning and Inference, 131, 63-88.

Feng, X. (2006). Confidence Intervals for Proportions With Focus on the US National Health and Nutrition Examination Survey. *Mater Thesis, Simon Fraiser University.*

Hall, P. (1982). Improving the Normal Approximation With Constructing One-sided ConfidenceIntervals for Binomial or Poisson Parameters. Biometrika, 69, 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, 24, 193-201.

Kott, P. S., Anderson, P. G. and Nerman O. (2001). Two-Sided Coverage Intervals for Small Proportion Based on Survey Data. Presented at Federal Committee on Statistical Methodology Research Confence, Washington, DC.

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.

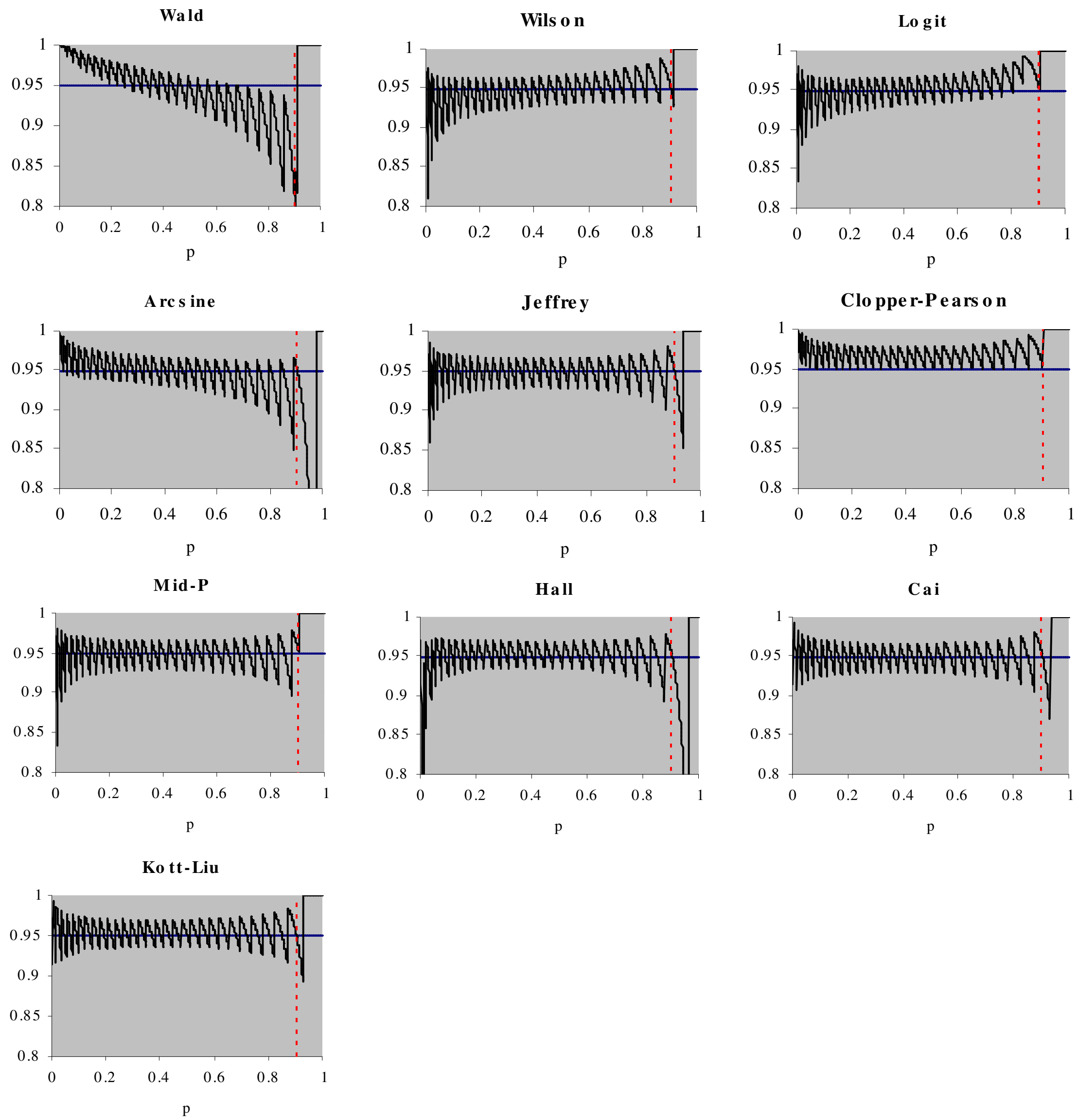


*Figure 1. Coverage Probabilities of Lower Bound at 95% Nominal Level: Simple Random Sample with n=30*

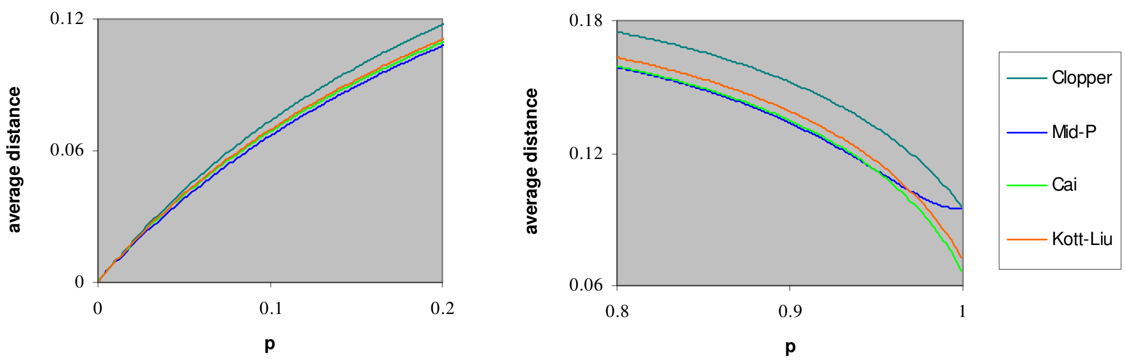


*Figure 2. Average Distance of Lower Bound at 95% Nominal Level: Simple Random Sample with n=30*

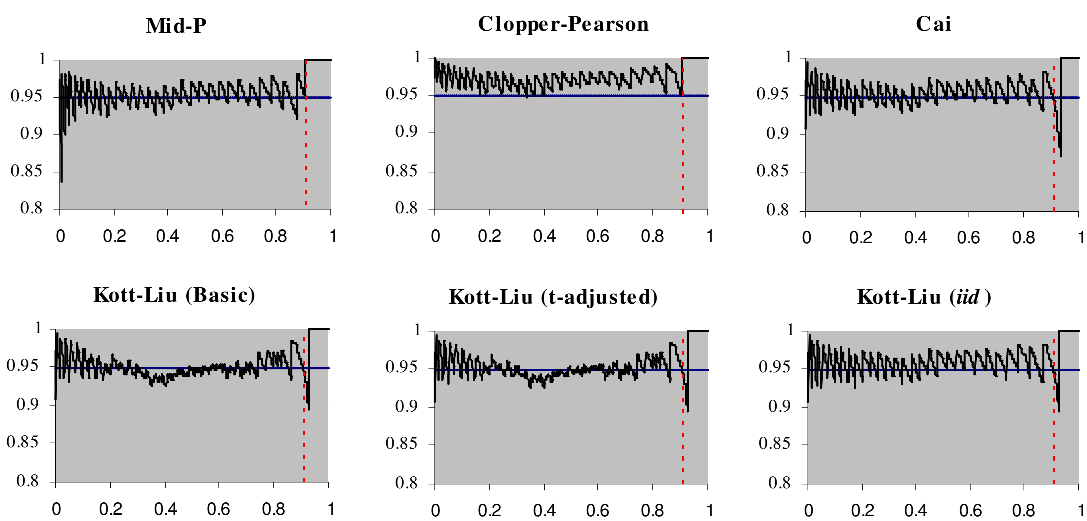


*Figure 3. Coverage Probabilities of Lower Bound at 95% Nominal Level for Setting B: Stratified Random Sample with* $n_1 = n_2 = n_3 = 10$; $p_1 = p - \Delta$, $p_2 = p$, $p_3 = p + \Delta$; $\Delta = p(1-p)$

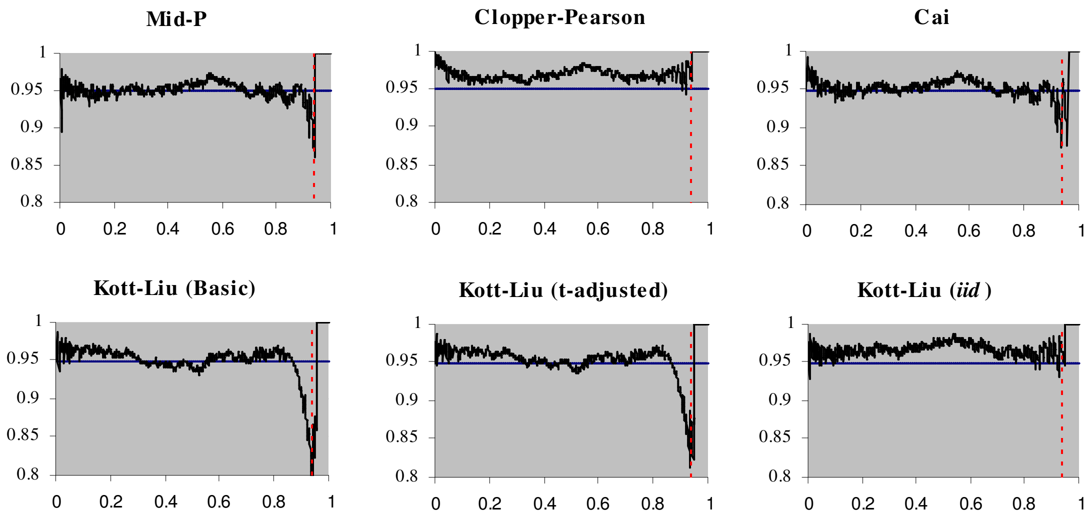


*Figure 4. Coverage Probabilities of Lower Bound at 95% Nominal Level for Setting D: Stratified Random Sample with $n_1 = n_3 = 10$, $n_2 = 30$; $p_1 = p - \Delta$, $p_2 = p$, $p_3 = p + \Delta$; $\Delta = p(1-p)$*

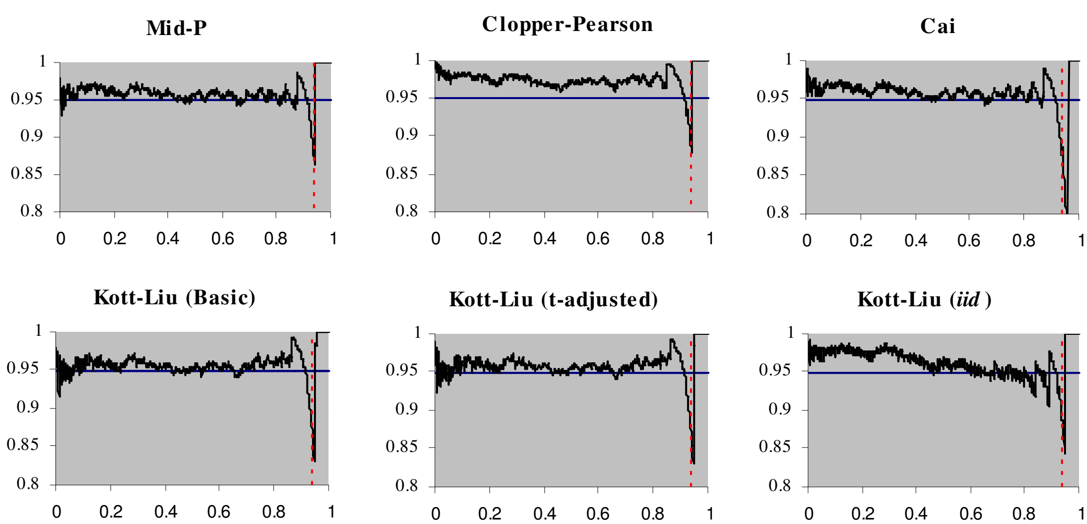


*Figure 5. Coverage Probabilities of Lower Bound at 95% Nominal Level for Setting E: Stratified Random Sample with $n_1 = n_2 = 10$, $n_3 = 30$; $p_1 = p - \Delta$, $p_2 = p$, $p_3 = p + \Delta$; $\Delta = p(1-p)$*

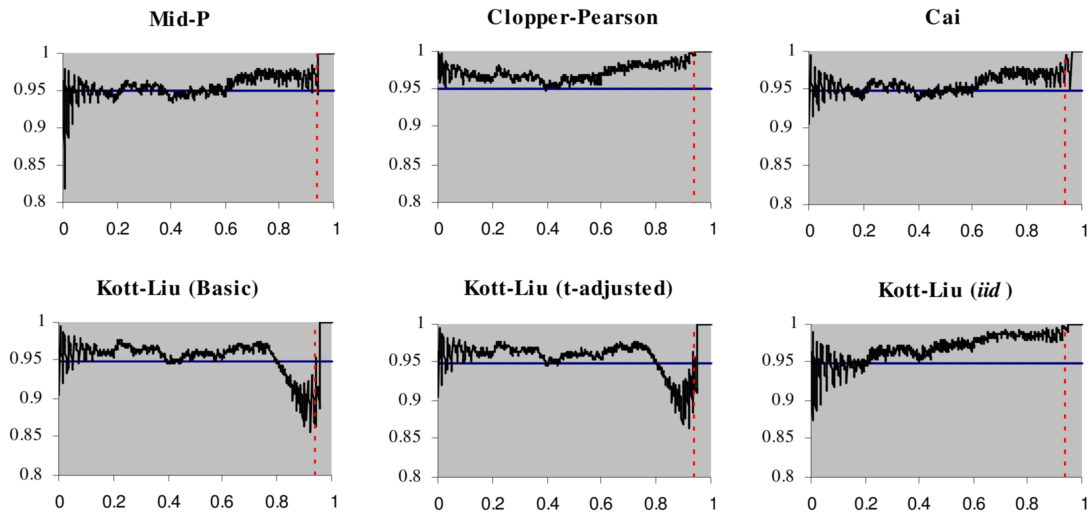


*Figure 6. Coverage Probabilities of Lower Bound at 95% Nominal Level for Setting F: Stratified Random Sample with $n_2 = n_3 = 10$, $n_1 = 30$; $p_1 = p - \Delta$, $p_2 = p$, $p_3 = p + \Delta$; $\Delta = p(1-p)$*

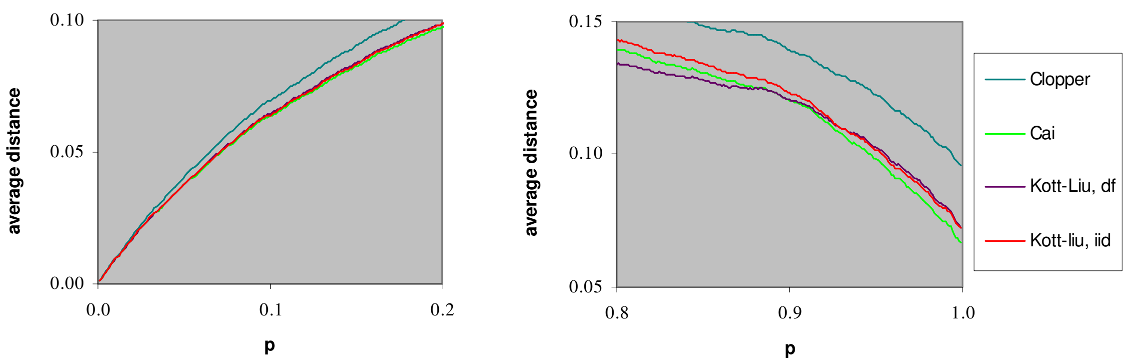


*Figure 7. Average Distances of Lower Bound at 95% Nominal Level for Setting B: Stratified Random Sample with $n_1 = n_2 = n_3 = 10$; $p_1 = p - \Delta$, $p_2 = p$, $p_3 = p + \Delta$; $\Delta = p(1-p)$*

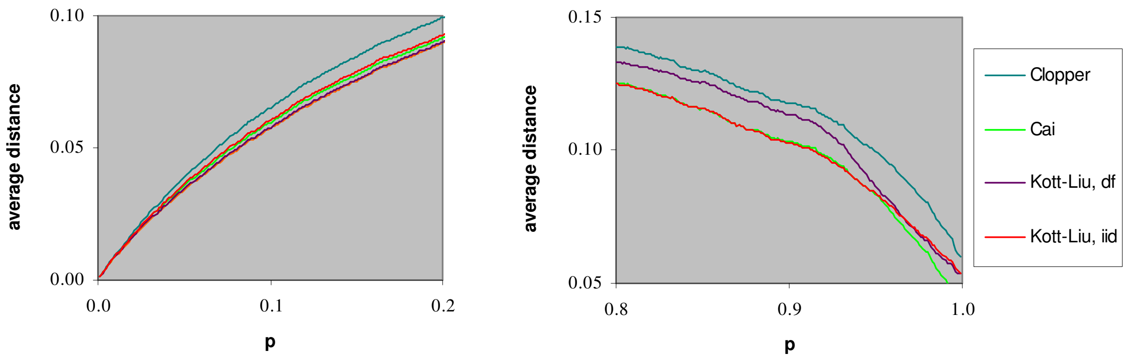


*Figure 8. Average Distances of Lower Bound at 95% Nominal Level for Setting C: Stratified Random Sample with $n_1 = n_3 = 10$, $n_2 = 30$; $p_1 = p_2 = p_3 = p$*