The Effect of Dependence on Confidence Intervals for a Population Proportion

Weiwen MIAO and Joseph L. GASTWIRTH

The binomial model is widely used in statistical applications. Usually, the success probability, p, and its associated confidence interval are estimated from a random sample. Thus, the observations are independent and identically distributed. Motivated by a legal case where some grand jurors could serve a second year, this article shows that when the observations are dependent, even slightly, the coverage probabilities of the usual confidence intervals can deviate noticeably from their nominal level. Several modified confidence intervals that incorporate the dependence structure are proposed and examined. Our results show that the modified Wilson, Agresti-Coull, and Jeffreys confidence intervals perform well and can be recommended for general use.

KEY WORDS: Coverage probability; Dependent observations; Expected length of confidence interval; Jury discrimination.

1. INTRODUCTION

Let $X_1, X_2, \ldots, X_n$ be binomial random variables with common unknown success probability p. Let n_i be the number of trials for each X_i and $N = n_1 + n_2 + \cdots + n_n$. When the X_i's are independent, $S_N = X_1 + X_2 + \cdots + X_n$ follows a binomial distribution and most textbooks estimate p by $\hat{p} = \frac{S_N}{N}$ and use the standard Wald confidence interval: $\hat{p} \pm z^* \sqrt{\hat{p}(1-\hat{p})/N}$, where z^* is the $100(1-\alpha/2)$th percentile of the standard normal distribution. It has been pointed out that the coverage probabilities of the standard Wald interval can be erratically poor even when p is not near 0 or 1 and alternative confidence intervals have been suggested. See, for example, Ghosh (1979), Blyth and Still (1983), Vollset (1993), Agresti and Coull (1998), Newcombe (1998), Santner (1998), and Chernick and Liu (2002). Brown, Cai, and DasGupta (2001) reviewed the performance of the standard Wald interval together with several alternatives and recommended the Wilson or equal-tailed Jeffreys prior interval for small samples ($N \leq 40$), and the Agresti-Coull interval for large samples ($N > 40$). They also provided a theoretical justification for their recommendations using asymptotic expansions of the coverage probabilities and expected length of these confidence intervals (Brown, Cai, and DasGupta 2002).

When the binomial random variables X_i are *dependent*, the distribution of S_N no longer is binomial. The effect of dependence on the sign test was examined by Wolff, Gastwirth, and Rubin (1967) and Gastwirth and Rubin (1971). Serfling (1968) showed that the level of the two-sample Wilcoxon test was similarly increased by positive correlation. Bernoulli trials with a Markov dependence have been used to model process arising in rainfall (Klotz 1973), metallurgy (Johnson and Klotz 1974), telecommunications (Crow 1979), meteorology (Katz 1981), and linguistics (Brainerd and Chang 1982). For first-order Markov dependent Bernoulli trials with $p = P(X_i = 1)$ and $\lambda = P(X_i = 1 | X_{i-1} = 1)$, $i = 1, 2, \ldots, n$, Ladd (1975) provided an algorithm for the confidence interval for the success probability p when the dependence parameter λ is known. Crow (1979) and Crow and Miles (1979) studied five approximate confidence intervals for the success probability p when the nuisance parameter λ is estimated. Bedrick and Aragon (1989) showed that the best power-divergence confidence intervals for p improve on those of Crow (1979).

Although the effect of the substantial degree of dependence arising in time-series and Markov dependent Bernoulli trials has been studied, there is little literature on the effect of a small degree of dependence on the properties of confidence intervals for p. Motivated by a legal case where the grand juries in consecutive years were dependent, this article studies the effect of small amount of dependence on the confidence intervals for a population proportion. It will be seen that all the usual confidence intervals are noticeably affected by the dependence that occurs *only* between consecutive observations. Modifications of the commonly used confidence intervals that incorporate the dependence are developed and studied. Our results show that the modified Wilson, Agresti-Coull, and Jeffreys confidence intervals have coverage probability close to the nominal level for most values of p, and their expected lengths are shorter than the other modified intervals. Hence they can be recommended for general use.

This article is organized as follows: Section 2 describes the actual legal case that motivated this research. The usual confidence intervals for a binomial proportion as well as the modified intervals that incorporate the dependence structure are introduced in Section 3. Section 4 describes examples of dependent binomial data. The motivating example is revisited in Section 4, using the modified confidence intervals constructed in the previous section. Section 5 compares the coverage probabilities of those intervals introduced in Section 3. A comparison of expected lengths of various intervals is provided in Section 6. Section 7 summarizes the results and makes recommendations based on them.

Weiwen Miao is Assistant Professor, Department of Mathematics and Computer Science, Macalester College, St. Paul, MN 55105 (E-mail: miao@macalester.edu). Joseph L. Gastwirth is Professor, Department of Statistics, George Washington University, Washington, DC 20052 (E-mail: jlgast@gwu.edu) and Guest Researcher at the Division of Cancer, Epidemiology and Genetics at National Cancer Institute. This research was supported in part by grant SES-0317956 from the National Science Foundation. The authors appreciate the helpful suggestions from the editor, associate editor, and both reviewers that substantially improved the article.

© 2004 American Statistical Association DOI: 10.1198/000313004343303
Material may be protected by copyright law (Title 17, U.S. Code)

2. THE MOTIVATING EXAMPLE

An actual legal case concerning possible discrimination against blacks in the selection of jurors serving on *grand juries* brought the problem to our attention. In many, but not all, jurisdictions, before an individual accused of a crime is brought to trial, the government first needs to convince a *grand jury* that there is sufficient evidence to justify filing a formal charge of the crime. The prosecutors do not need to convince the grand jury that the accused is guilty "beyond a reasonable doubt" as is required at trial, rather, only the lesser standard of "probable cause" needs to be shown. If the grand jury agrees that the evidence provided by the state meets this threshold, the accused is indicted by the grand jury and the trial process initiated. The composition of grand juries, like trial juries, should reflect the demographics of the area. In *Moultrie v. Martin* (4th Cir. 1982, 690 F.2d 1078), the petitioner was arrested in September 1977 for the murder of the county sheriff. A grand jury composed of 3 blacks and 15 whites returned an indictment against him. At the subsequent trial, the petitioner moved to quash the indictment on the grounds that he was denied "equal protection" because blacks were under-represented on the grand jury that indicted him. But the judge allowed the trial to proceed. After his conviction, the defendant appealed to the federal Court of Appeals for the Fourth Circuit. This court stated that the claim was based entirely on statistics. Data concerning the demographic mix of grand juries for the 1971–1977 period was considered as the court was examining whether the *process* of selecting grand juries was fair. Courts recognize that even in a fair system, all subgroups of the population may not be represented in proportion to their fractions of eligible individuals residing in the relevant area on a particular jury of 12 people or grand jury of 18.

In South Carolina, where the *Moultrie* case occurred, grand juries are chosen from the same master list as trial juries; however, their length of service is different. Each year, 12 individuals from the master list are chosen at random for service on the county's grand jury. For the next year, 6 of those 12 are randomly selected to remain on the grand jury for one additional year and 12 new people, randomly selected from the master list, are added. Thus, a total of 18 people serve on the county's sole grand jury for each year. Because of this overlap in service of 6 panel members, the racial compositions of grand juries in consecutive years are *not* statistically independent. The actual data from the *Moultrie* case is given in Table 1. Note that only the racial mix of the grand jury for each year was submitted into evidence. More detailed data reporting both the number of blacks among the 12 new grand jurors chosen from the master list and the number of blacks among the six "holdover" grand jurors was not presented.

The above process differs from the usual one in which each jury or grand jury is an independent random sample from the eligible population. To describe it precisely, let X_i be the number of black grand jurors in the ith year, $i = 1, 2, \ldots, n$, then $S_N = X_1 + \ldots, X_n$ is total number of black grand jurors in all n years, counting the duplicates. Let p be the true proportion of blacks eligible for grand jury service in the county. The problem of interest is whether the confidence interval for p, constructed using the data reported in Table 1, contains the true value of the black proportion of eligible grand jurors. This true value of p is usually determined from external data, such as the most recent Census or voter rolls, when they are considered representative. In the *Moultrie* case, the court observed that 1977 blacks formed 38% of the voting rolls and accepted 0.38 as the true value of p for the entire period. Consequently, we are interested in whether the confidence interval for p contains 0.38 or not.

Table 1. Number of Blacks on Grand Juries of 18 in Colleton County

Year	1971	1972	1973	1974	1974	1976	1977
# of Blacks	1	5	5	7	7	4	3

In the usual situation where there is no overlap between consecutive juries or grand juries, the X_i's are *independent* binomial random variables with success probability p. Then S_N has a binomial distribution and the usual confidence intervals for p can be used. However, S_N in the *Moultrie* case no longer has a binomial distribution because of the carry-over grand jury selection process. The appellate opinion actually noted the problem, but the judges accepted a test of the hypothesis of whether $p = 0.38$, using the usual independence assumption as they are not statisticians. Using the fact that X_i's are 1-dependent binomial random variables, Gastwirth and Miao (2002) showed that the effect of dependence on the p value of testing $p = 0.38$ versus $p \neq 0.38$ is quite noticeable. For example, the continuity corrected p value for the period 1972–1977 changed from 0.059 when the dependence is ignored to 0.129 when the dependence is taken into consideration.

One might try to avoid the complications arising from dependence by restricting the statistical analysis to the individuals selected each year from the master list. Ignoring the grand jurors carried over for a second year, however, would allow a county to intentionally discriminate against blacks in the carry over process, while appointing the minimal number of blacks in the new grand jurors needed to "pass" the standard binomial test. This is a general problem occurring when one is concerned with two-stage employment practices, as shown in Gastwirth (1997) who studied methods for assessing the fairness of both hiring and promotion practices.

3. THE CONFIDENCE INTERVALS

This section reviews the usual confidence intervals for a binomial proportion p and describes modified intervals that incorporate the dependence structure. Let $X_1, \ldots X_n$ be binomial random variables with common success probability p. When the X_i's are *independent*, the S_N follows a binomial distribution with parameters $N = n_1 + \cdots + n_n$ and p. When X_i's are dependent, the corresponding confidence intervals rely on the structure of the dependence. When S_N has an asymptotic normal distribution, confidence intervals are created by using the variance that accounts for the dependence in place of $Np(1-p)$, the variance of S_N in the independent case. In other words, one defines N^* by $N^*p(1-p) = \text{var}(S_N)$, where $\text{var}(S_N)$ is the variance of S_N for the dependent data. A confidence interval appropriate for independent data is then modified for the dependent setting by replacing N by N^*.

Material may be protected by copyright law (Title 17, U.S. Code)

3.1 A Review of Confidence Intervals For Independent Data

Before proceeding to the dependent case, it is helpful to review some of the commonly used confidence intervals for *independent* data. Let $\hat{p} = \frac{S_N}{N}$ and z^* be the $100(1-\alpha/2)$th percentile of the standard normal. The standard Wald ($\mathrm{CI_S}$), Wilson ($\mathrm{CI_W}$), and Agresti-Coull ($\mathrm{CI_{AC}}$) intervals are:

$$\mathrm{CI_S} = \hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{N}},$$

$$\mathrm{CI_W} = \frac{S_N + \frac{1}{2}z^{*2}}{N + z^{*2}} \pm \frac{z^* N^{\frac{1}{2}}}{N + z^{*2}} \sqrt{\hat{p}(1-\hat{p}) + \frac{z^{*2}}{4N}},$$

and

$$\mathrm{CI_{AC}} = \tilde{p} \pm z^* \sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{N}}}$$

with

$$\tilde{N} = N + z^{*2},\ \tilde{p} = \frac{S_N + \frac{1}{2}z^{*2}}{N + z^{*2}}.$$

with lower (upper) bound equals 0 (1) when $S_N = 0$ (N)

Notice that the Wilson interval is obtained by solving $|\hat{p} - p| \le z^* \sqrt{\frac{p(1-p)}{N}}$ for p. The Agresti-Coull interval has the same simple form as the Wald interval, but it replaces $\hat{p}$ by $\tilde{p}$, the center of the Wilson interval. Following Fleiss (1981), Bohning (1994), and Newcombe (1998), the continuity corrected Wilson interval is obtained by solving $|\hat{p} - p| - \frac{1}{2N} \le z^* \sqrt{\frac{p(1-p)}{N}}$ for p, that is,

$$\mathrm{CI_{ccW}} = \frac{(2N\hat{p} + z^{*2} \pm 1) \pm z^* \sqrt{z^{*2} \pm (2 \mp \frac{1}{N}) + 4N\hat{p}(1-\hat{p}) \mp 4\hat{p}}}{2(N + z^{*2})}$$

with the lower (upper) bound set to be 0 (1) if $S_t = 0$ (N).

Following Brown et al. (2001), the Jeffreys prior interval ($\mathrm{CI_J}$) is:

$$[B(\alpha/2, S_N + 1/2, N - S_N + 1/2),\ B(1-\alpha/2, S_N + 1/2, N - S_N + 1/2]$$

with lower (upper) bound equals 0 (1) when $S_N = 0$ (N) and $B(\alpha; m_1, m_2)$ denotes the αth quantile of a Beta(m_1, m_2) distribution.

The Clopper-Pearson ($\mathrm{CI_{CP}}$) exact interval is:

$$\mathrm{CI_{CP}} = [B(\alpha/2, S_N, N - S_N + 1),\ B(1-\alpha/2, S_N + 1, N - S_N)]$$

with the lower (upper) bound set to be 0 (1) if $S_N = 0$ (N).

3.2 Confidence Intervals for Dependent Data

Note that $E(S_N) = p\sum n_i = Np$ and $\mathrm{var}(S_N) = \sum_{i=1}^{n} \mathrm{var}(X_i) + 2\sum_{i<j} \mathrm{cov}(X_i, X_j) = Np(1-p) + 2p(1-p)\sum_{i<j}\sqrt{n_i n_j}\rho_{ij} = p(1-p)[N + 2\sum_{i<j}\sqrt{n_i n_j}\rho_{ij}]$, where ρ_{ij} is the correlation coefficient between X_i and X_j. When $\{X_i\}$ satisfy conditions ensuring that S_N is asymptotically normal, such as m-dependence (Hoeffding and Robbin 1948) or various mixing conditions (Gastwirth and Rubin 1975; Herrndorf 1984; Rosenblatt 1984), then

$$\left| \frac{S_N - Np}{\sqrt{p(1-p)\left[N + 2\sum_{i<j}\sqrt{n_i n_j}\rho_{ij}\right]}} \right| = \left| \frac{\hat{p} - p}{\sqrt{p(1-p)\left[\frac{1}{N} + \frac{2}{N^2}\sum_{i<j}\sqrt{n_i n_j}\rho_{ij}\right]}} \right| = \frac{\hat{p} - p}{\sqrt{p(1-p)/N^*}} \Rightarrow N(0,1),$$

where $N^* = [\frac{1}{N} + \frac{2}{N^2}\sum_{i<j}\sqrt{n_i n_j}\rho_{ij}]^{-1}$. This is precisely the formula for the asymptotic distribution of $\hat{p}$ in the *independent* case with N replaced by N^*. This relationship suggests the following modification of the confidence intervals for dependent data. The modified Wald confidence interval ($\mathrm{CI_S^d}$) is

$$\mathrm{CI_S^d} = \hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{N^*}}.$$

The modified Wilson interval ($\mathrm{CI_W^d}$) is obtained by solving

$$\left| \frac{\hat{p} - p}{\sqrt{p(1-p)/N^*}} \right| \le z^*$$

for p, that is,

$$\mathrm{CI_W^d} = \frac{\hat{p}N^* + \frac{1}{2}z^{*2}}{N^* + z^{*2}} \pm \frac{z^* N^{*\frac{1}{2}}}{N^* + z^{*2}} \sqrt{\hat{p}(1-\hat{p}) + \frac{z^{*2}}{4N^*}}.$$

The modified Agresti-Coull interval is obtained by using the center of the modified Wilson interval and the same form of the modified Wald interval:

$$\mathrm{CI_{AC}^d} = \tilde{p}^* \pm z^* \sqrt{\frac{\tilde{p}^*(1-\tilde{p}^*)}{\tilde{N}^*}}$$

with

$$\tilde{N}^* = N^* + z^{*2},\ \tilde{p}^* = \frac{\hat{p}N^* + \frac{1}{2}z^{*2}}{N^* + z^{*2}}.$$

Solving the inequality $|\hat{p} - p| - \frac{1}{2N} \le z^*\sqrt{p(1-p)/N^*}$ for p yields the modified continuity corrected Wilson interval:

$$\mathrm{CI_{ccW}^d} = \frac{(2N^*\hat{p} + z^{*2} \pm \frac{N^*}{N})}{2(N^* + z^{*2})} \pm \frac{z^*\sqrt{z^{*2} + 4N^*\hat{p}(1-\hat{p}) - \frac{N^*}{N^2} \pm \frac{2N^*}{N} \mp \frac{4N^*\hat{p}}{N}}}{2(N^* + z^{*2})}$$

with lower (upper) bound equals 0 (1) when $S_N = 0$ (N).

The Jeffreys ($\mathrm{CI_J^d}$) and the Clopper-Pearson confidence intervals ($\mathrm{CI_{CP}^d}$) for dependent data are also obtained by replacing N by N^* and S_N by $\hat{p}N^*$. That is,

$$\mathrm{CI_J^d} = [B(\alpha/2, \hat{p}N^* + 1/2, N^* - \hat{p}N^* + 1/2),\ B(1-\alpha/2, \hat{p}N^* + 1/2, N^* - \hat{p}N^* + 1/2]$$

Material may be protected by copyright law (Title 17, U.S. Code)

$$\mathrm{CI}^d_{\mathrm{CP}} = [B(\alpha/2, \hat{p}N^*, N^* - \hat{p}N^* + 1),$$
$$B(1-\alpha/2, \hat{p}N^* + 1, N^* - \hat{p}N^*)]$$

with lower (upper) bound equal to 0 (1) when $S_N = 0$ (N).

4. APPLICATIONS TO DEPENDENT DATA

Some examples of dependent binomial data are described in this section. To illustrate and compare the methods, the modified confidence intervals are constructed for the data from the *Moultrie* case.

4.1 Examples of Dependent Data

1. When the $X_1, \ldots, X_n$ are *m-dependent*, the central limit theorem for m-dependent random variables (Hoeffding and Robbin 1948) implies that S_N converges to a normal distribution and the above formulas apply. In the *Moultrie v. Martin* case, X_i are 1-dependent with $n_i = 18$, $\rho_{ij} = 1/3$ if $|j-i| = 1$ and 0 otherwise (Gastwirth and Miao 2002). Consequently, $N = 18n$ and $N^* = [\frac{1}{18n} + \frac{2}{(18n)^2} \cdot 18 \cdot (n-1) \cdot \frac{1}{3}]^{-1} = \frac{(18n)^2}{30n-12}$, which is less than N as long as $n > 1$. Indeed, $N^*/N = \frac{18n}{30n-12} < 0.7$ when $n > 2$.

2. Let $Y_1, \ldots, Y_n$ be a strictly stationary first-order autoregressive process with independent errors, or equivalently, $Y_{i+1} = \rho Y_i + \epsilon_i$, $i = 1, 2, \ldots, n$, where the ϵ_i are iid random variables symmetric about 0. Let

$$X_i = \begin{cases} 1 & Y_i \geq 0 \\ 0 & Y_i < 0, \end{cases} \qquad \text{and} \qquad p = P(X_i = 1).$$

Then X_i are *dependent* binomial random variables with $n_i = 1$ and $N = n$. Let $S_N = X_1 + \cdots + X_N$. The S_N is the sign test statistic used to test whether the Y_is have median 0. It has been shown (Wolff, Gastwirth, and Rubin 1967) that $\mathrm{var}(S_N) = \sum_{k=-N}^{N} (N - |k|)\,\rho_k = Np(1-p) + 2\sum_{k=1}^{N}(N-k)\rho_k$, where $\rho_k = \mathrm{cov}(X_1, X_k)$. Thus, if S_N is asymptotically normal, then the formulas in Section 3.2 apply with $N^* = \frac{p(1-p)N^2}{\mathrm{var}(S_N)} = \frac{p(1-p)N^2}{Np(1-p)+2\sum_{k=1}^{N}(N-k)\rho_k}$.

2(a). In the special case of a first-order Gaussian process where the Y_i's are standard normals, $p = 0.5$. Wolff et al. (1967) proved that $\mathrm{var}(S_N) = \frac{1}{4}N + \frac{1}{\pi}\sum_{k=1}^{N}(N-k)\arcsin\rho^k$ and S_N is asymptotically normal. Consequently, $N^* = \frac{N^2}{N + \frac{4}{\pi}\sum_{k=1}^{N}(N-k)\arcsin\rho^k}$.

2(b). In the corresponding first-order autoregressive double exponential process, the Y_i's have double exponential distributions with mean 0 and variance 2. Wolff et al. (1967) showed that S_N is asymptotically normally distributed with $E(S_N) = N/2$ and $\mathrm{var}(S_N) = \frac{1}{4}\left[N\frac{1+\rho}{1-\rho} - \frac{2\rho(1-\rho^N)}{(1-\rho)^2}\right]$. In this case, $N^* = \frac{N^2(1-\rho)^2}{N(1-\rho^2)-2\rho(1-\rho^N)}$. Note that even for first-order autoregressive process with the same ρ, the variance of S_N and the N^* depend on the stationary distribution, that is, the sign test is no longer distribution free.

3. Consider Bernoulli trials with first-order Markov dependence. Klotz (1973) used this model for measurable precipitation in the month of June at Madison, WI. Let $X_1, \ldots, X_n$ be dependent Bernoulli random variables with $P(X_i = 1) = p$ and $P(X_{i+1} = 1 | X_i = 1) = \lambda$, $i = 1, 2, \ldots, n$. Then $n_i = 1$ and $N = n$. Let $\rho = (\lambda - p)/(1-p)$ be the correlation coefficient between X_i and X_{i+1}. Klotz (1973) showed that $S_N = X_1 + \ldots X_N$ is asymptotically normal and the variance of S_N is $\mathrm{var}(S_N) = p(1-p)[N + \frac{2N\rho}{1-\rho} - \frac{2\rho(1-\rho^N)}{(1-\rho)^2}]$. Consequently, when the parameter ρ is known, the $N^* = [\frac{1}{N} + \frac{2\rho}{N(1-\rho)} - \frac{2\rho(1-\rho^N)}{N^2(1-\rho)^2}]^{-1}$. Usually one would estimate ρ using the estimated p.

4.2 Reanalysis of the *Moultrie v. Martin* Data

Recall that each year's grand jury consists of 18 people: 12 are randomly selected from the population and 6 are randomly selected from the 12 chosen as new grand jurors in the previous year. Let X_i be the number of black grand jurors for the ith year. Gastwirth and Miao (2002) showed that the correlation coefficient between X_i and X_j is $1/3$ if $|j-i| = 1$ and 0 otherwise. Because there was only one black grand juror in the first year, 1971, noticeably less than the data for other years, the court decided to disregard that year's data. Considering the remaining six years, 1972–1977, $N = 18 * 6 = 108$ while $N^* = [\frac{1}{108} + \frac{2}{108^2}(6-1) * 18 * \frac{1}{3}]^{-1} = 69.4$. There were 31 black grand jurors (counting duplicates) for the time period 1972–1977, so $S_N = 31$. The corresponding 95% confidence intervals are given in Table 2. For comparative purposes, the confidence intervals that treat the X_i's as independent binomial random variables are given in Table 2.

Table 2 clearly indicates that the confidence intervals incorporating dependence are wider than those that incorrectly assumed independence. Actually the ratio of the length of modified confidence interval to that of the corresponding independent confidence interval is about 1.25, which is approximately $\sqrt{N/N^*}$. More importantly, as 0.38 was accepted as the true black proportion for the entire period, all the modified confidence intervals include 0.38, while only two of the independent confidence intervals do. They were the continuity corrected Wilson and Clopper-Pearson intervals, which are known to be conservative (Brown et al. 2001). The *continuity corrected* Wald interval (assuming independence) was (0.180, 0.394), which includes 0.38. This is consistent with the p value of 0.059 of the test of $p = 0.38$ noted in Section 2. As a referee pointed out the effect of the dependence, which reduces the effective sample size from N to N^*, is the same for both hypothesis tests and confidence intervals, preserving the usual correspondence between them.

Table 2. Confidence Intervals for the Moultrie Data

Confidence interval	Assuming independence	Incorporating dependence
Standard Wald	(0.202,0.372)	(0.181,0.393)
Wilson	(0.210,0.379)	(0.194,0.402)
Agresti-Coull	(0.210,0.379)	(0.193,0.403)
Continuity Corrected Wilson	(0.206,0.383)	(0.190,0.407)
Jeffrey	(0.208,0.377)	(0.191,0.401)
Clopper-Pearson	(0.204,0.382)	(0.185,0.408)

5. COVERAGE PROBABILITIES

The coverage probability of any of these confidence intervals

Material may be protected by copyright law (Title 17, U.S. Code)

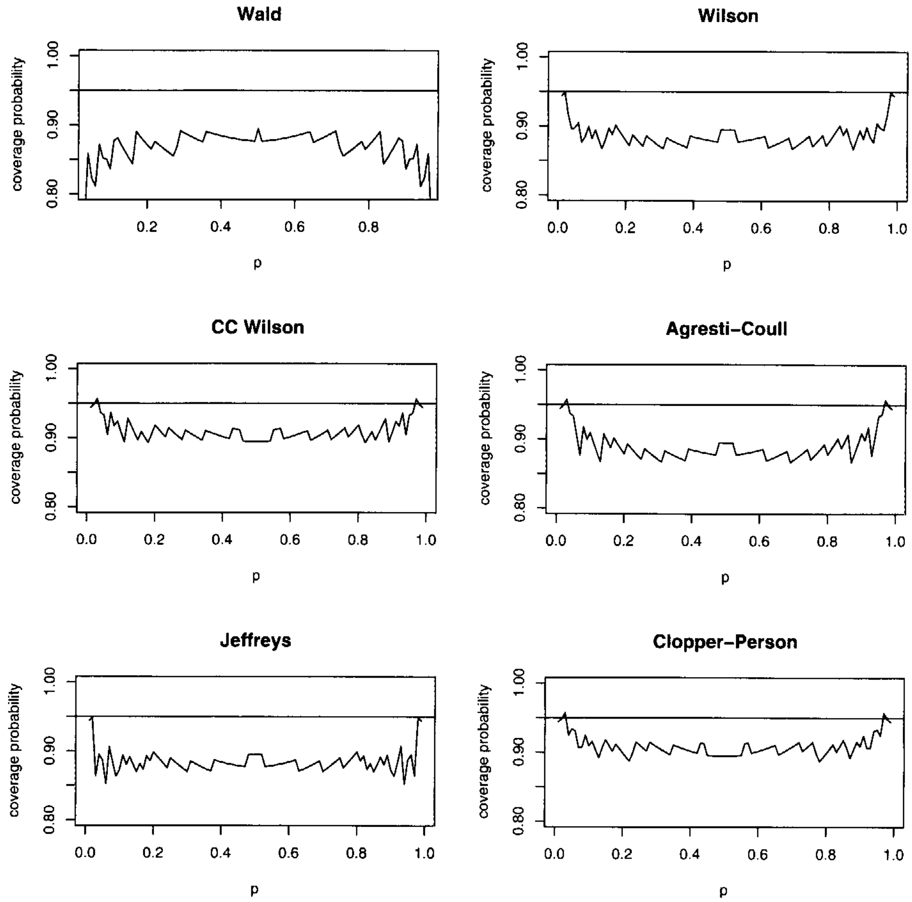

Figure 1. Coverage probability for confidence intervals ignoring *dependence. (f = 12, c = 6, n = 6, ρ = 1/3).*

is given by:

$$\begin{aligned} \text{CP} &= P(p \in \text{CI}) = \sum_{k=0}^{N} P(p \in \text{CI} | S_N = k) \cdot P(S_N = k) \\ &= \sum_{k=0}^{N} I(k,p) P(S_N = k), \\ I(k,p) &= \begin{cases} 1, & \text{if} \quad p \in \text{CI} \quad \text{when} \quad S_N = k \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

This section examines the coverage probabilities of the confidence intervals for a slightly more generalized model of the grand jury selection process. Each year let f be the number of first-time grand jurors randomly selected from the population (in *Moultrie*, $f = 12$). Let c ($c \leq f$) be the carry-over grand jurors, selected randomly from the previous jurors (in *Moultrie*, $c = 6$). Thus, there are $f + c$ grand jurors each year. Over n years, there are $N = (f+c)n$ grand jurors. Gastwirth and Miao (2002) showed that the density function of $S_N = X_1 + \ldots X_n$ is

$$P(S_N = i) = \sum_{k=\max(0, i-(n-1)(f+c))}^{\min(i, f+c)} \left[\binom{f+c}{k} p^k (1-p)^{(f+c)-k} \cdot \Sigma_{b_1 + \cdots + b_{n-1} = i-k} \left(a_{b_1} \ldots a_{b_{n-1}} \right) \right] \quad (1)$$

with $a_i = \Sigma^i_{k \geq i/2} \binom{c}{i-k} \binom{f-c}{2k-i} p^k (1-p)^{f-k}$, $i = 0, 1, 2, \ldots, f + c$.

Equation (1) enables us to calculate the exact coverage probabilities for all the intervals considered here. First, we explore the effect of ignoring the dependence on the usual confidence intervals. Gastwirth and Miao (2002) showed that the correlation, ρ, of black grand jurors in *consecutive* years is $\frac{c}{f+c}$. For nonconsecutive years, $\rho = 0$. Notice that $\rho \leq 0.5$ as $c \leq f$.

Figure 1 presents the exact coverage probabilities for a nominal 95% confidence interval for each of the six methods described in Section 3.1, where the dependence between consecutive years is ignored. The graph, as well as those appearing later in the article, were produced using R software with p from $(0, 1)$ with increments of 0.01. We took $n = 6$, $f = 12$ and $c = 6$ as in the *Moultrie* case. Then $N = 18 * 6 = 108, \rho = \frac{1}{3}$, and $N^* = 69.4$. Clearly all six intervals have coverage probability much lower than the nominal 0.95 level for most values of p. These findings confirm the importance of considering the dependence structure in analyzing data. Similar results were obtained for the cases $n = 6, f = 17, c = 1, \rho = \frac{1}{18}$, and $n = 6, f = c = 9, \rho = \frac{1}{2}$. Consistent with the formula of var(S_N), as the positive ρ_{ij} increase, the dependence between the observations increases, N^* decreases and the coverage probabilities of all the confidence intervals ignoring dependence decline.

Figure 2 presents the exact coverage probabilities for the modified confidence intervals incorporating the dependence structure when $\rho = 1/3$. The modified Wald interval still has coverage

Material may be protected by copyright law (Title 17, U.S. Code)

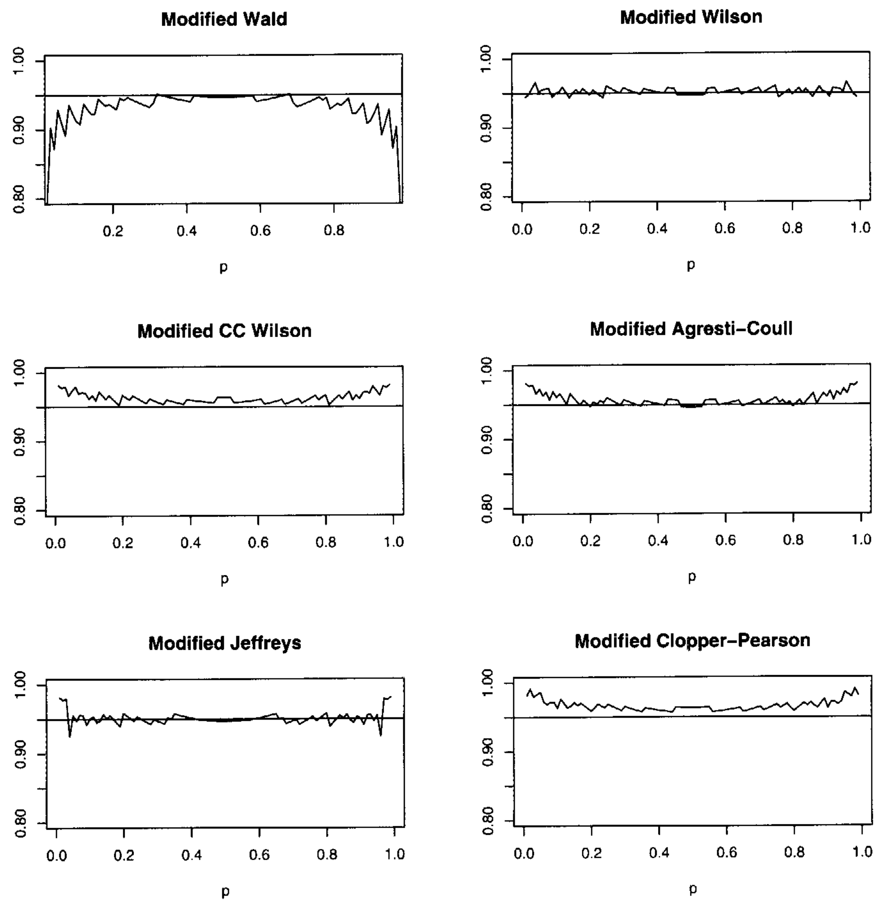

Figure 2. Coverage probability for modified confidence intervals. (f = 12, c = 6, n = 6, ρ = 1/3).

probabilities lower than the nominal 0.95 for most values of p, especially those near 0 or 1. These results are consistent with those of Brown et al. (2001) in the independent case and use of that method is not recommended. The modified Wilson and Jeffreys intervals have coverage probability near the nominal 0.95 for all values of p. The coverage probability for modified Agresti-Coull interval is close to the nominal 0.95 for most values of p. Only when p is near 0 or 1, its coverage probabilities are higher than the nominal 0.95. The modified continuity corrected Wilson and Clopper-Pearson intervals are conservative with coverage probability higher than 0.95 for all values of p. These findings agree with those of Brown et al. (2001) in the case of independent binomial data as can be seen by comparing Figure 2 with their Figure 5.

The amount of dependence for the grand jury selection process is relatively small even when $\rho = 1/2$, the largest possible value, as the number of black grand jurors in *nonconsecutive* years are independent. To explore a process with a greater degree of dependence, we examined the behavior of the sign test on the first-order autoregressive Gaussian process in Example 2(a) of Section 4. We simulated the process $Y_{i+1} = \rho Y_i + \epsilon_i, i = 1, 2, \ldots n$, where Y_1 is a standard normal random variable and the ϵ_i are iid $N(0, 1 - \rho^2)$. Using the simulated number of Y_i's > 0, we constructed 95% confidence intervals for $p = P(Y_i > 0) = 0.5$ and checked whether these intervals included 0.5. As expected, all confidence intervals that *incorrectly* assumed independence had coverage probabilities far from the nominal 0.95 even for sample sizes as large as 200. Also, the higher the value of $|\rho|$, the further the coverage probability differed from 0.95. The modified confidence intervals performed much better compared to those that were not modified. However, when $n = 50$ and $\rho > 0.8$, the modified Wald interval had coverage less than 0.85, while the continuity corrected Wilson and Clopper-Pearson intervals had coverage higher than 0.975. The modified Wilson, Agresti-Coull, and Jeffreys interval performed well when N^* is not small ($N^* > 10$). When N^* is small, for example in the case when $n = 30, \rho = 0.9, N^*$ is only 3.14, the coverage probabilities for the modified Wilson, Agresti-Coull, and Jeffreys intervals are equal to 1. This indicates that further research needs to be done to obtain efficient confidence intervals when there is substantial dependence among the observations.

6. THE EXPECTED LENGTHS OF CONFIDENCE INTERVALS THAT INCORPORATE DEPENDENCE

The expected length (W) of a confidence interval is given by

$$E(W) = \sum_{k=0}^{N} (U(k, f, c, n) - L(k, f, c, n))\, P(S_N = k),$$

where U and L are the upper and lower bounds of the confidence intervals. Using the density function of S_N given in (1) and the lower and upper bounds of each type of confidence interval given in Section 3, $E(W)$ can be calculated exactly. Here, we consider only the expected length of the modified Wilson, continuity corrected Wilson, Agresti-Coull, Jeffreys, and Clopper-Pearson intervals which have desirable coverage probabilities when the dependence effect is included.

Material may be protected by copyright law (Title 17, U.S. Code)

We looked at the expected length of those intervals for $N = 108$ and $\rho = \frac{1}{18}, \frac{1}{3}$, and $\frac{1}{2}$, respectively. For each type of interval, the greater the dependence, the longer the expected width. For example, when the true population proportion $p = 0.38$, the expected lengths for the modified Wilson interval are 0.187, 0.221, and 0.239 for $\rho = \frac{1}{18}, \frac{1}{3}$, and $\frac{1}{2}$, respectively. This is not surprising as the variance of $\hat{p}$ increases with $\rho = \frac{c}{f+c}$ as n and $(f + c)$ are fixed. For the same degree of dependence, the expected lengths of the modified Clopper-Pearson and continuity corrected Wilson intervals are longer than those of the modified Wilson, Agresti-Coull, and Jeffreys intervals. This is consistent with the earlier finding that their coverage probabilities are higher than the other three confidence intervals. The expected lengths for the modified Wilson, Agresti-Coull, and Jeffreys intervals are about the same.

7. CONCLUSION AND RECOMMENDATION

The degree of dependence arising in the hold-over grand jury data is far less than that occurring in the typical time series data as the grand jurors in *nonconsecutive* years are actually *independent*. Nevertheless, even with a small amount of dependence, the coverage probabilities of all the usual confidence intervals were seen to be far less than the nominal 0.95 value in Section 5. Thus, any confidence interval needs to use the correct variance of S_N which incorporates the dependence structure of the data. These results reinforce van Belle's (2002) observation that ignoring dependence is a major pitfall in data analysis.

Our results indicate when one has dependent binomial data with success probability p, the modified Wilson, Agresti-Coull, and Jeffreys intervals have coverage probability close to the nominal level for most values of p and their expected lengths are shorter than the other modified intervals. Hence those intervals can be recommended for dependent data. Further research is needed to obtain appropriate modified confidence intervals for processes with long-range dependence (Beran 1994) as well as random-effects models where X_i are independent binomial with parameters n_i and p_i and the p_i come from a distribution.

[Received January 2003. Revised March 2004.]

REFERENCES

Agresti, A., and Coull, B. A. (1998), "Approximate is Better Than 'Exact' for Interval Estimation of Binomial Proportions," *The American Statistician*, 52, 119–126.

Bedrick, E. J., and Aragon, J. (1989), "Approximate Confidence Intervals for the Parameters of a Stationary Binary Markov Chain," *Technometrics*, 31, 437–448.

Beran, J. (1994), *Statistics for Long-Memory Processes*, New York: Chapman and Hall.

Blyth, C. R., and Still, H. A. (1983), "Binomial Confidence Intervals," *Journal of American Statistical Association*, 78, 108–116.

Bohning, D. (1994), "Better Approximate Confidence Intervals for a Binomial Parameter," *The Canadian Journal of Statistics*, 22, 207–218.

Brainerd, B., and Chang, S. M. (1982), "Number of Occurrences in Two-State Markov Chains, With an Application in Linguistics," *The Canadian Journal of Statistics*, 10, 225–231.

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

——— (2002), "Confidence Intervals for a Binomial Proportion and Asymptotic Expansions," *The Annals of Statistics*, 30, 160–201.

Chernick, M., and Liu, C. Y. (2002), "The Saw-Toothed Behavior of Power Versus Sample Size and Software Solutions: Single Binomial Proportion Using Exact Methods," *The American Statistician*, 56, 149–155.

Crow, E. L. (1979), "Approximate Confidence Intervals for a Proportion from Markov Dependent Trials," *Communications in Statistics—Simulation and Computation*, B8, 1–42.

Crow, E. L., and Miles, M. J. (1979), "Validation of Estimators of a Proportion from Markov Dependent Trials," *Communication of Statistics—Simulation and Computation*, B8, 25–52.

Fleiss, J. L. (1981), *Statistical Methods for Rates and Proportions* (2nd ed.), New York: Wiley.

Gastwirth, J. L. (1997), "Statistical Evidence in Discrimination Cases," *Journal of the Royal Statistical Society*, Series A, 160, 289–303.

Gastwirth, J. L., and Miao, W. (2002), "The Potential Effect of Statistical Dependence in the Analysis of Data in Jury Discrimination Cases: *Moultrie v. Martin* Reconsidered," *Jurimetrics*, 43, 115–128.

Gastwirth, J. L., and Rubin, H. (1971), "Effect of Dependence on the Level of Some One-Sample Tests," *Journal of the American Statistical Association*, 66, 816–820.

——— (1975), "The Asymptotic Distribution Theory of the Empirical CDF for Mixing Stochastic Processes," *The Annals of Statistics*, 3, 809–824.

Ghosh, B. K. (1979), "A Comparison of Some Approximate Confidence Intervals for the Binomial Parameter," *Journal of American Statistical Association*, 74, 894–900.

Herrndorf, N. (1984), "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables," *The Annals of Probability*, 12, 141–153.

Hoeffding, W., and Robbins, H. E. (1948), "The Central Limit Theory for Dependent Random Variables," *Duke Mathematics Journal*, 15, 773–780.

Johnson, C. A., and Klotz, J. H. (1974), "The Atom Probe and Markov Chain Statistics of Clustering," *Technometrics*, 16, 483–493.

Katz, R. W. (1981), "On Some Criteria for Estimating the Order of a Markov Chain," *Technometrics*, 23, 243–249.

Klotz, J. H. (1973), "Statistical Inference in Bernoulli Trials with Dependence," *The Annals of Statistics*, 373–379.

Ladd, D. W. (1975), "An Algorithm for the Binomial Distirbution with Dependent Trials," *Journal of the American Statistical Association*, 70, 333–340.

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

Rosenblatt, M. (1984), "Asymptotic Normality, Strong Mixing and Spectral Density Estimate," *The Annals of Probability*, 12, 1167–1180.

Santner, T. J. (1998), "Teaching Large-Sample Binomial Confidence Intervals," *Teaching Statistics*, 20, 20–23.

Serfling, R. J. (1968), "The Wilcoxon Two-Sample Statistics on Strong-Mixing Processes," *Annals of Mathematical Statistics*, 1202–1209.

van Belle, G. (2002), *Statistical Rules of Thumb*, New York: Wiley.

Vollset, S. E. (1993), "Confidence Intervals for a Binomial Proportion," *Statistics in Medicine*, 12, 809–824.

Wolff, S. S., Gastwirth, J. L., and Rubin, H. (1967), "The Effect of Autoregressive Dependence on a Nonparametric Test," *IEEE Transaction on Information Theory*, 13, 311–313.

Material may be protected by copyright law (Title 17, U.S. Code)