SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853-7501

TECHNICAL REPORT NO. 795

May 1988
Revised April 1989

**INTERVAL ESTIMATION OF THE MARGINAL
PROBABILITY OF SUCCESS FOR THE
BETA-BINOMIAL DISTRIBUTION**

by

David Wypij¹ and Thomas J. Santner²

¹Department of Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115.

²School of Operations Research and Industrial Engineering, Upson Hall, Cornell University, Ithaca, NY 14853-7501.

ABSTRACT

Correlated binomial outcomes occur in many areas such as teratology and marketing. We study the problem of confidence interval estimation of the common marginal probability of success for correlated binary observations, focusing on the beta-binomial model. New conservative small sample confidence intervals are proposed for this problem based on extensions of the Tail and Sterne-Crow intervals for independent binomial responses. Some properties of these intervals are derived, comparisons with previously proposed methods are given, and recommendations regarding their use are made.

Key Words Beta-binomial model; marginal probability of success; confidence interval estimation; Tail intervals; Sterne-Crow intervals.

1. INTRODUCTION

Recently there has been renewed interest in the development of large and small sample confidence intervals for the success probability of binary data. For example, Blyth and Still (1983) and Casella (1986) discuss both cases for independent and identically distributed (iid) Bernoulli trials. Duffy and Santner (1986) and the references therein construct small sample intervals for iid Bernoulli trials based on a multi-stage stopping rule. Crow and Miles (1979) determine approximate intervals for data following a stationary first-order binary Markov chain.

This paper proposes small sample confidence intervals for the (marginal) success probability when T mutually independent vectors $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})$, $1 \leq i \leq T$, are observed from a beta-binomial distribution. This one-sample beta-binomial model and its extension to the general regression case have been applied in the study of chromosomes (Skellam, 1948), market research (Chatfield and Goodhardt, 1970), disease incidence (Griffiths, 1973), teratology (Williams, 1975), and mutagenesis (Otake and Prentice, 1984). The remainder of this section introduces notation and motivates the developments of the following sections by displaying the achieved coverage probability of several approximate confidence intervals.

The beta-binomial distribution can be obtained by drawing iid Bernoulli trials $Y_{i1}, \dots, Y_{in_i}$ with success probability p_i after p_i is drawn from the beta distribution

$$\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p_i^{\alpha-1} (1 - p_i)^{\beta-1}, \quad 0 < p_i < 1,$$

for $1 \leq i \leq T$, where $\alpha > 0$, $\beta > 0$. It is convenient to reparametrize to the marginal probability of success

$$\pi = P[Y_{ij} = 1] = \alpha / (\alpha + \beta) \in (0, 1)$$

and the within-group correlation

$$\rho = \text{Corr}(Y_{i1}, Y_{i2}) = 1/(\alpha + \beta + 1) \in (0,1).$$

The probability mass function of Y_i is

$$f_i(y_{i1}, \dots, y_{in_i}) = \binom{n_i}{y_i} \frac{\prod_{j=0}^{y_i-1} (\pi + \gamma j) \prod_{j=0}^{n_i-y_i-1} (1 - \pi + \gamma j)}{\prod_{j=0}^{n_i-1} (1 + \gamma j)} \quad (1.1)$$

for $y_i \in \{0, 1, \dots, n_i\}$, where $y_i = \sum_{j=1}^{n_i} y_{ij}$, $\sum_{j=0}^x q_j = 1$ for $x < 0$, and $\gamma = \rho/(1 - \rho)$. Thus $(Y_1, \dots, Y_T)$ is a sufficient statistic for the beta-binomial family indexed by parameters

$(\pi, \rho) \in (0,1) \times (0,1)$ where $Y_i = \sum_{j=1}^{n_i} Y_{ij}$, $1 \leq i \leq T$. Prentice (1986) notes that (1.1)

extends smoothly to the boundary of the parameter space $(\pi, \rho) \in [0,1] \times [0,1]$. For example, the distribution corresponding to correlation $\rho = 1$ is

$$f_i(y_{i1}, \dots, y_{in_i}) = \begin{cases} \pi^{I[y_i = n_i]} (1 - \pi)^{1 - I[y_i = n_i]}, & y_i = 0 \text{ or } n_i \\ 0 & , \text{ otherwise.} \end{cases}$$

Thus $Y_{i1} = \dots = Y_{in_i} = 1$ ($Y_i = n_i$) with probability π and $Y_{i1} = \dots = Y_{in_i} = 0$ ($Y_i = 0$) with probability $1 - \pi$.

One approach to forming approximate π confidence intervals is to estimate π and form an interval based on the (normalized) estimator's asymptotic distribution. A second method is to invert a family of large sample tests of $H_0: \pi = \pi_0$ versus $H_A: \pi \neq \pi_0$ (as π_0 ranges over $[0,1]$).

The first of three examples of the former is Williams (1975) who estimates π by maximum likelihood and its standard error by the appropriate element of the inverse of the sample information based on the beta-binomial likelihood

$$L(\pi, \rho) = \prod_{i=1}^T f_i(y_{i1}, \dots, y_{in_i}). \quad (1.2)$$

Gladden (1979) uses the overall proportion of successes

$$\hat{\pi}_G = Y_+ / n_+,$$

with the jackknife estimate of standard error

$$\hat{\sigma}(\hat{\pi}_G) = \left[\frac{T-1}{T} \sum_{i=1}^T (X_i - \bar{X})^2 \right]^{1/2}$$

where $Y_+ = \sum_{i=1}^T Y_i$, $X_i = (Y_i - n_i \hat{\pi}_G) / (n_+ - n_i)$, $n_+ = \sum_{i=1}^T n_i$, and $\bar{X} = \frac{1}{T} \sum_{i=1}^T X_i$. She proves that

$$P[\pi \in \hat{\pi}_G \pm z_{\alpha/2} \hat{\sigma}(\hat{\pi}_G)] \rightarrow 1 - \alpha \quad (1.3)$$

as $T \rightarrow \infty$ where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution.

Southward and Van Ryzin (1972) and Van Ryzin (1975) suggest using the average proportion of successes

$$\hat{\pi}_S = \frac{1}{T} \sum_{i=1}^T \frac{Y_i}{n_i}$$

with

$$\hat{\sigma}(\hat{\pi}_S) = \left[\frac{1}{T(T-1)} \sum_{i=1}^T \left[\frac{Y_i}{n_i} - \hat{\pi}_S \right]^2 \right]^{1/2}$$

and prove

$$P[\pi \in \hat{\pi}_S \pm z_{\alpha/2} \hat{\sigma}(\hat{\pi}_S)] \rightarrow 1 - \alpha \quad (1.4)$$

as $T \rightarrow \infty$ under very weak conditions which include the beta-binomial model. Both of the latter two approaches are nonparametric in that (1.3) and (1.4) hold under H_0 for mixing distributions satisfying mild conditions. The intervals are identical when $n_1 = \dots = n_T$.

The likelihood ratio (LR) test is recommended by many authors as the test of $H_0: \pi = \pi_0$ versus $H_A: \pi \neq \pi_0$ and for forming confidence intervals by test inversion (e.g., Cox and Hinkley, 1974). The corresponding acceptance region is

$$A(\pi_0) = \{y: 2 \ln\left(\frac{L(\hat{\pi}, \hat{\rho})}{L(\pi_0, \hat{\rho}_0)}\right) \leq \chi_{1, \alpha}^2\}$$

where $(\hat{\pi}, \hat{\rho})$ is the joint maximum likelihood estimator (MLE) of (π, ρ) , $\hat{\rho}_0$ is the restricted MLE of ρ when $\pi = \pi_0$, $\ln(\cdot)$ denotes natural logarithm, and $\chi_{1, \alpha}^2$ is the upper α -quantile of the χ_1^2 distribution. The LR test confidence set contains the global MLE $\hat{\pi}$, and lies in $[0, 1]$ although it need not be a symmetric interval about $\hat{\pi}$.

Figure 1 plots the achieved coverage probabilities as a function of π for the Gladen, Southward–Van Ryzin, and LR 90% nominal intervals for beta-binomial data with $\rho = 0.5$, $T = 10$ groups of size $n_i = 3$, $1 \leq i \leq 10$. It shows that the approximate confidence intervals do not achieve the desired coverage levels and can perform quite poorly for certain π values. The undercoverage is more pronounced for more extreme values of ρ . Thus alternative methods are needed to guarantee the desired coverage probabilities for small or moderately sized correlated binary data sets.

Sections 2 and 3 propose new conservative small sample intervals for π based on extensions of the Tail and Sterne–Crow intervals for the independent binomial case. The final section illustrates the intervals on teratology data.

2. TAIL CONFIDENCE INTERVALS

Let $\mathcal{Y} = \{\underline{y} = (y_1, \dots, y_T): 0 \leq y_i \leq n_i \text{ for } 1 \leq i \leq T\}$ be the sample space for the sufficient statistic $\underline{Y} = (Y_1, \dots, Y_T)$ of the beta-binomial data and let $\underline{n} = (n_1, \dots, n_T)$. The key idea of the Tail method is to invert a family of acceptance regions $A(\pi_0)$ of $H_0: \pi = \pi_0$ versus $H_A: \pi \neq \pi_0$ which are defined as the complement of a union of two "extreme" regions each having null probability $\leq \alpha/2$. Since this is a nuisance parameter problem, test sizes must be calculated using

$$Q(E|\pi) = \sup_{0 \leq \rho \leq 1} P[\underline{Y} \in E | \pi, \rho] \quad (2.1)$$

for $E \subset \mathcal{Y}$. Appendix 1 proves Theorem 2.1, namely that $Q(E|\pi)$ is monotone in π for "tail events" E .

THEOREM 2.1: $Q(E|\pi)$ is nonincreasing in π for any "lower tail" event of the form $E = \{\underline{y} \in \mathcal{Y}: y_+ \leq j\}$ with $0 \leq j \leq n_+$. Conversely, $Q(E|\pi)$ is nondecreasing in π for any "upper tail" event of the form $E = \{\underline{y} \in \mathcal{Y}: y_+ \geq j\}$ with $0 \leq j \leq n_+$.

For $\alpha < 1$, level $100(1 - \alpha)\%$ Tail intervals are defined by inverting the family of acceptance regions

$$A(\pi_0) = \mathcal{Y} \setminus (\underline{R}_0 \cup \bar{R}_0)$$

of $H_0: \pi = \pi_0$ versus $H_A: \pi \neq \pi_0$ where

$$\bar{R}_0 = \{\underline{y} \in \mathcal{Y}: Q([Y_+ \geq y_+]) | \pi_0) \leq \alpha/2\}$$

and

$$\underline{R}_0 = \{\underline{y} \in \mathcal{Y}: Q([Y_+ \leq y_+]) | \pi_0) \leq \alpha/2\}.$$

The set $A(\pi_0)$ is a nonempty $100(1 - \alpha)\%$ acceptance region of the form

$$\{\underline{y} \in \mathcal{Y}: L(\pi_0) \leq y_+ \leq U(\pi_0)\} \quad (2.2)$$

for some $L(\pi_0) \leq U(\pi_0)$. To see this, observe that $\underline{R}_0$ and $\bar{R}_0$ must be disjoint since otherwise

$$\begin{aligned} 1 &= P[\underline{R}_0 \cup \bar{R}_0 | \pi_0, \rho] \\ &\leq P[\underline{R}_0 | \pi_0, \rho] + P[\bar{R}_0 | \pi_0, \rho] \leq \alpha/2 + \alpha/2 < 1. \end{aligned}$$

for all ρ which is a contradiction. Thus $P[A(\pi_0) | \pi_0, \rho] \geq 1 - \alpha$ for all ρ follows from the definition of $\underline{R}_0$ and $\bar{R}_0$. Theorem 2.2 gives an explicit characterization of the Tail limits.

THEOREM 2.2: For any $\underline{y} \in \mathcal{Y}$, $\{\pi_0: \underline{y} \in A(\pi_0)\} = [\underline{\pi}(y_+), \bar{\pi}(y_+)]$ where $\underline{\pi}(0) = 0$, $\bar{\pi}(n_+) = 1$, and otherwise are the solutions of

$$Q([Y_+ \geq j] | \underline{\pi}(j)) = \alpha/2 \text{ and } Q([Y_+ \leq j] | \bar{\pi}(j)) = \alpha/2. \quad (2.3)$$

PROOF: Fix $0 < j < n_+$. Then by Theorem 2.1,

$$\begin{aligned} \underline{\pi}(j) \leq \pi_0 \leq \bar{\pi}(j) &\Leftrightarrow \begin{cases} Q([Y_+ \geq j] | \pi_0) \geq Q([Y_+ \geq j] | \underline{\pi}(j)) = \alpha/2 \\ \text{and} \\ Q([Y_+ \leq j] | \pi_0) \leq Q([Y_+ \leq j] | \bar{\pi}(j)) = \alpha/2 \end{cases} \\ &\Leftrightarrow L(\pi_0) \leq j \leq U(\pi_0) \\ &\Leftrightarrow j \in A(\pi_0). \end{aligned}$$

The cases $j = 0$ and n_+ follow by similar arguments. $\square$

REMARK 2.1: Characterization (2.3) can be solved in closed form for $j = 0$ and n_+ but requires iterative calculation in all other cases. For any ρ ,

$$\begin{aligned} P[Y_+ \leq 0 | \pi, \rho] &= P[Y_1 = \dots = Y_T = 0 | \pi, \rho] \\ &= (1 - \pi)^T \prod_{i=1}^T \prod_{j=1}^{n_i-1} \frac{1 - \pi + \gamma j}{1 + \gamma j} \end{aligned}$$

where $\gamma = \rho/(1 - \rho)$. This expression attains its supremum over γ as $\gamma \rightarrow \infty$ ($\rho \rightarrow 1$) so

$$Q([Y_+ \leq 0] | \pi) = (1 - \pi)^T \stackrel{\text{set}}{=} \alpha/2 \Rightarrow \bar{\pi}(0) = 1 - (\alpha/2)^{1/T}.$$

Analogous arguments give $\underline{\pi}(n_+) = (\alpha/2)^{1/T}$. $\square$

Tail intervals have the following symmetry and monotonicity properties:

- P1 Equivariance under relabeling of successes and failures: if $y_{ij} \rightarrow 1 - y_{ij}$ for all i and j then $(\underline{\pi}, \bar{\pi}) \rightarrow (1 - \bar{\pi}, 1 - \underline{\pi})$.
- P2 Symmetry about $1/2$ when n_+ is even and $y_+ = n_+/2$: $\underline{\pi}(n_+/2) = 1 - \bar{\pi}(n_+/2)$.
- P3 Monotonicity in y_+ : for fixed T , $\underline{n}$, and α , $\bar{\pi}(j) \leq \bar{\pi}(j+1)$ and $\underline{\pi}(j) \leq \underline{\pi}(j+1)$.
- P4 Monotonicity in n_i : for fixed T , α , and $j \geq 0$, $\bar{\pi}(j)$ and $\underline{\pi}(j)$ are nonincreasing in each n_i .
- P5 Monotonicity in T : for fixed α and $j \geq 0$, if additional failures are added in new groups then $\bar{\pi}(j)$ and $\underline{\pi}(j)$ are nonincreasing.
- P6 Monotonicity in α : for fixed T , $\underline{n}$, and $j \geq 0$, $\bar{\pi}(j)$ is nonincreasing and $\underline{\pi}(j)$ is nondecreasing in α .

Each of the properties P1 – P6 has an intuitive interpretation. For example, P3 says that when a success is observed in place of a failure the interval estimate of π should not be revised downward. Properties P4 and P5 say that when additional failures are observed the interval should not be revised upward.

From P3 – P5 it follows that when additional successes are observed in existing or new groups the interval should not be revised downward. A consequence of P6 is that Tail intervals become wider as the coverage probability increases.

The proof of P2 follows from P1; the proofs of the remaining properties are similar to one another and only that for P1 is given below. Wypij (1987) proves the remaining properties.

THEOREM 2.3: Property P1 holds for Tail intervals for any α , $\underline{n}$, and $\underline{y}$.

PROOF: Equivariance holds for $j = 0$ and n_+ by Remark 2.1. For any $\underline{y} \in \mathcal{Y}$ let $\underline{n} - \underline{y} = (n_1 - y_1, \dots, n_T - y_T)$. Then for any π and ρ ,

$$P[\underline{Y} = \underline{y} | \pi, \rho] = P[\underline{Y} = \underline{n} - \underline{y} | 1 - \pi, \rho] \quad (2.4)$$

from (1.1). Thus

$$P[Y_+ \geq j | \pi, \rho] = P[Y_+ \leq n_+ - j | 1 - \pi, \rho].$$

Hence for any $0 < j < n_+$,

$$\alpha/2 = Q([Y_+ \geq j] | \pi(j)) = Q([Y_+ \leq n_+ - j] | 1 - \pi(j))$$

which implies

$$1 - \pi(j) = \bar{\pi}(n_+ - j). \quad \square$$

When $n_1 = \dots = n_T = 1$, the proposed intervals reduce to the Clopper–Pearson (1934) Tail intervals for independent Bernoulli trials. In general, the beta–binomial Tail limits are more conservative than the Clopper–Pearson limits. Formally this means that for fixed T , n , $0 \leq j \leq n_+$, and α , the beta–binomial interval for $y_+ = j$ contains the Clopper–Pearson interval based on j successes out of n_+ independent trials. To see this note that

$$\alpha/2 = Q([Y_+ \leq j] | \bar{\pi}(j)) \geq P[Y_+ \leq j | \bar{\pi}(j), \rho = 0]$$

and the right hand probability is nonincreasing in the success probability.

Despite their attractive properties, Tail intervals have two practical difficulties. First, their calculation requires the numerical solution of

$$\sup_{0 \leq \rho \leq 1} P[E | \pi, \rho] = \alpha/2 \quad (2.5)$$

involving a supremum for an appropriate set E . Wypij and Santner (1989) discuss numerical issues that arise in the solution of (2.5) for fixed π . More importantly, tail intervals can be extremely conservative as Figure 2 illustrates. While the degree of conservatism can be decreased (see Section 4), this characteristic is of sufficient concern to motivate the study of less conservative but more computationally intensive Sterne–Crow (S/C) intervals.

3. STERNE–CROW INTERVALS

The idea of S/C is to generate shorter intervals by using smaller acceptance regions. Systems with shorter intervals have closer-to-nominal coverage probabilities. However, additional computational effort is required to determine smaller acceptance regions. Sterne

(1954), Crow (1956), Blyth and Still (1983), and Casella (1986) determine minimal size acceptance regions for the case of iid Bernoulli trials.

The algorithm below partitions $[0, 1/2]$ by $0 = \pi_0 < \dots < \pi_M < 1/2$ and constructs a $100(1 - \alpha)\%$ acceptance region for $H_0: \pi = \pi_i$ versus $H_A: \pi \neq \pi_i$ of the form

$$A(\pi_i) = \{\underline{y} \in \mathcal{Y}: L_i \leq y_+ \leq U_i\} \quad (3.1)$$

where $0 \leq L_i \leq U_i \leq n_+$. The partition mesh determines the accuracy of the resulting confidence limits. The algorithm moves $A(\pi_{i+1})$ to the "right" of $A(\pi_i)$ in the sense that $L_i \leq L_{i+1}$ and $U_i \leq U_{i+1}$. Acceptance regions are formed for $\pi_i > 1/2$ by symmetry, thus insuring the resulting confidence intervals are equivariant (Property P1).

Let

$$q(E|\pi) = \inf_{0 \leq \rho \leq 1} P[\underline{Y} \in E | \pi, \rho]$$

for $E \subset \mathcal{Y}$ and $C[L, U] = \{\underline{y} \in \mathcal{Y}: L \leq y_+ \leq U\}$.

Step 0 [Initialization]

- 0.1 Construct a grid $0 = \pi_0 < \pi_1 < \dots < \pi_M < 1/2$.
- 0.2 Set $L_0 = 0$, $U_0 = 0$, $A(\pi_0) = C[L_0, U_0]$, and $i = 0$.
- 0.3 Go to Step 1.

Step 1 [Induction Step]

- 1.1 If $i = M$, go to Step 6. Otherwise set $L = L_i$ and $U = U_i$.
- 1.2 If $q(C[L, U] | \pi_{i+1}) \geq 1 - \alpha$ then go to Step 2. Otherwise go to Step 3.

Step 2 [Elimination]

- 2.1 If $L + U = n_+$, go to Step 5.
- 2.2 If $q(C[L + 1, U] | \pi_{i+1}) \geq 1 - \alpha$ then set $L = L + 1$ and go to Step 2. Otherwise go to Step 4.

Step 3 [Addition]

- 3.1 Set $U = U + 1$.
- 3.2 If $q(C[L, U] | \pi_{i+1}) \geq 1 - \alpha$ then go to Step 2. Otherwise go to Step 3.

Step 4 [Alternate Acceptance Regions]

- 4.1 If $L + U - 2 > n_+$, go to Step 5.
- 4.2 If $q(C[L + 1, U + 1] | \pi_{i+1}) \geq 1 - \alpha$ and $\Delta(L, U | \pi_{i+1}) \geq \Delta(L + 1, U + 1 | \pi_{i+1})$, where $\Delta(\cdot)$ is defined by (3.2) below, then set $L = L + 1$, $U = U + 1$, and go to Step 2. Otherwise go to Step 5.

Step 5 [Acceptance Region]

- 5.1 Set $i = i + 1$, $A(\pi_i) = C[L, U]$, $L_i = L$, $U_i = U$, and go to Step 1.

Step 6 [Inversion]

- 6.1 For $i = 0(1)M$ set $\pi_{2M+1-i} = 1 - \pi_i$, $L_{2M+1-i} = n_+ - U_i$, and $U_{2M+1-i} = n_+ - L_i$.
- 6.2 For $j = 0(1)n_+$, set $\underline{\pi}(j) = \min \{\pi_i : U_i \geq j\}$, and $\bar{\pi}(j) = \max \{\pi_i : L_i \leq j\}$.

Step 4.2 chooses between same-sized $100(1 - \alpha)\%$ acceptance regions, selecting the region which is closest to being "equal tailed." To accomplish this we propose using

$$\Delta(L, U | \pi) = \sum_{j=0}^B |P[Y_+ < L | \pi, \rho_j] - P[Y_+ > U | \pi, \rho_j]| \quad (3.2)$$

for a grid of values $0 \leq \rho_0 < \dots < \rho_B$ where $|\cdot|$ denotes absolute value. The function $\Delta(\cdot)$ measures the difference between the tail probabilities of the rejection region averaged over the $\{\rho_j\}$ values. Other measures of tail inequality also make sense; for example one could use

$$\Delta(L, U | \pi) = \int |P[Y_+ < L | \pi, \rho] - P[Y_+ > U | \pi, \rho]| h(\rho) d(\rho)$$

where $h(\cdot)$ is a prior distribution for ρ . The example below uses (3.2) with $B = 10$ and $\rho_j = j/10$ for $0 \leq j \leq 10$. See Blyth and Still (1983) for a discussion of the analogous issue in the case of independent Bernoulli trials.

Each of the acceptance sets $\{A(\pi_i)\}_{i=1}^{2M+1}$ has coverage probability at least $1 - \alpha$. For any $0 \leq i \leq M$ and $0 \leq \rho \leq 1$,

$$P[Y \in A(\pi_i) | \pi_i, \rho] \geq 1 - \alpha$$

by construction. For $M + 1 \leq i \leq 2M + 1$, the symmetry property (2.4) yields

$$\inf_{0 \leq \rho \leq 1} P[Y \in A(\pi_i) | \pi_i, \rho] = \inf_{0 \leq \rho \leq 1} P[Y \in A(1 - \pi_i) | 1 - \pi_i, \rho] \geq 1 - \alpha$$

from the construction in Step 6.1 of the algorithm.

That the inversion of the $A(\pi_i)$ yields intervals and can be performed as in Step 6 results from the monotonicity of the L_i and U_i in π_i (see Blyth and Still (1983)). By construction the endpoints satisfy

$$L_i \leq L_{i+1} \text{ and } U_i \leq U_{i+1} \quad (3.3)$$

for $0 \leq i < M$ and $M + 1 \leq i < 2M + 1$. To show that (3.3) holds for the boundary case $i = M$ in which L_{i+1} is constructed from L_i requires analysis of the algorithm.

The following proof depends upon a lemma which is given in Appendix 2 and a generalization which is conjectured.

LEMMA 3.1: $L_M \leq L_{M+1}$ and $U_M \leq U_{M+1}$.

PROOF: It suffices to prove $L_M + U_M \leq n_+$ since then $L_M \leq L_{M+1} = n_+ - U_M$ and $U_M \leq U_{M+1} = n_+ - L_M$ where the equalities hold by construction. We show $L_j + U_j \leq n_+$ for $0 \leq j \leq M$ by induction on j . The case $j = 0$ is true by Step 0 of the algorithm. Suppose the claim holds for $0 \leq j < k \leq M$. Applying the algorithm in Step 1 with $i = k - 1$ gives the following. The only case that has to be analyzed is Step 3 since Steps 2.1 and 4.1 prevent $L_k + U_k > n_+$. In Step 3, $L = L_i$ and $U_i \leq n_i - L_i \Rightarrow$

$$q(C[L_i, n_i - L_i] | \pi_i) \geq q(C[L_i, U_i] | \pi_i) \geq 1 - \alpha. \quad (3.4)$$

Appendix 2 states that $q(C[L_i, n_i - L_i] | \pi)$ is nondecreasing in π over $[0, 1/2]$. Hence

$$q(C[L_i, n_i - L_i] | \pi_k) \geq 1 - \alpha$$

which shows U_k will never be moved past $n_i - L_i$ and completes the proof. $\square$

It can now be proved that Step 6.2 performs the proper inversion of the acceptance regions.

THEOREM 3.1: For any $0 \leq j \leq n_+$ and $0 \leq i \leq 2M + 1$, $\{\underline{y} \in \mathcal{Y} : y_+ = j\} \subset A(\pi_i)$
 $\Leftrightarrow \underline{\pi}(j) \leq \pi_i \leq \bar{\pi}(j)$.

PROOF: By Step 6.2, $\underline{\pi}(j) = \min \{\pi_k : U_k \geq j\}$ and $\bar{\pi}(j) = \max \{\pi_k : L_k \leq j\}$. By equation (3.3), L_i and U_i are nondecreasing in i . Hence $\{\pi_k : U_k \geq j\} = \{\pi_k : \underline{\pi}(j) \leq \pi_k\}$ and

$\{\pi_k: L_k \leq j\} = \{\pi_k: \pi_k \leq \bar{\pi}(j)\}$, and so

$$\{\underline{y} \in \mathcal{Y}: y_+ = j\} \subset A(\pi_i) \Leftrightarrow L_i \leq j \leq U_i \Leftrightarrow \underline{\pi}(j) \leq \pi_i \leq \bar{\pi}(j). \quad \square$$

The S/C intervals constructed by the algorithm above have the symmetry properties P1 – P3. However, unless additional conditions are imposed by the algorithm, counterexamples can be constructed which show that properties P4 – P6 can fail. The proofs of P1 – P3 require similar arguments and only P1 is proved below. Wypij (1987) proves the remaining properties.

THEOREM 3.2: Property P1 holds S/C intervals for any α , $\underline{n}$, and $\underline{y}$.

PROOF: For any $0 \leq j \leq n_+$ we have

$$\begin{aligned} \bar{\pi}(n_+ - j) &= \max \{\pi_i: L_i \leq n_+ - j\} \\ &= \max \{1 - \pi_i: L_{2M+1-i} \leq n_+ - j\} \\ &= 1 - \min \{\pi_i: L_{2M+1-i} \leq n_+ - j\} \\ &= 1 - \min \{\pi_i: n_+ - U_i \leq n_+ - j\} \\ &= 1 - \min \{\pi_i: U_i \geq j\} \\ &= 1 - \underline{\pi}(j). \quad \square \end{aligned}$$

When $n_1 = \dots = n_T = 1$, the proposed intervals reduce to a variant of the Sterne–Crow intervals for independent Bernoulli trials. In particular, the variant depends on the choice of the function $\Delta(\cdot)$; our choice of $\Delta(\cdot)$ reduces to the rule of minimizing the difference between the tail probabilities (Blyth and Still, 1983, p. 110). In general, the

S/C acceptance regions are more conservative than their independent counterparts. Formally this means that for $0 \leq i \leq 2M + 1$, the S/C acceptance region $A(\pi_i)$ is a $100(1 - \alpha)\%$ acceptance region for $H_0: \pi = \pi_i$ versus $H_A: \pi \neq \pi_i$ based on n_+ independent Bernoulli trials. To see this note that

$$1 - \alpha \leq q(A(\pi_i) | \pi_i) \leq P[\underline{Y} \in A(\pi_i) | \pi_i, \rho = 0].$$

4. DISCUSSION AND AN EXAMPLE

Tail intervals have monotonicity properties in $\underline{n}$, α , and T that are not automatically guaranteed by S/C intervals. Tail intervals can be used to form one-sided confidence bounds. In addition they have the computational advantage that when only the interval corresponding to the single outcome $y_+ = j$ is desired then (2.3) alone need be solved whereas S/C intervals require the computation of an entire family of confidence sets $A(\pi_i)$.

However, using symmetry properties of the S/C intervals can permit early stopping of the algorithm. For example, when $y_+ = j$ the algorithm can stop constructing $A(\pi_i)$ as soon as $\{\underline{y} \in \mathcal{Y}: y_+ = j\}$ is first added to and then eliminated from the sequence of acceptance regions.

S/C intervals have the advantage of superior operating characteristics. Figure 2 plots the achieved coverage probability for both systems of intervals when $T = 10$, $n_i = 3$ for $1 \leq i \leq T$, and $\alpha = 0.10$ when the true correlation is $\rho = 0.5$. Figure 3 plots the expected length of the intervals as a function of π for both systems under the same conditions.

Less conservative versions of both Tail and S/C intervals can be determined when the experimenter has prior knowledge about ρ . Suppose it is known that $\rho \leq \rho_{\max}$ (given). Then the function $q(E | \pi)$ can be replaced by

$$\inf_{0 \leq \rho \leq \rho_{\max}} P[E | \pi, \rho].$$

The improvement of making this change is illustrated in Figure 4 which plots the achieved coverage probability of the restricted range S/C interval ($\rho_{\max} = 0.5$) and the unrestricted range S/C interval when the true $\rho = 0.5$. Similar improvements occur in the expected length. Similarly, if a prior distribution $h(\cdot)$ is known for ρ then it is appropriate to use the average coverage $q(E|\pi) = \int P[Y \in E|\pi, \rho] h(\rho) d\rho$ in the algorithm. Wypij and Santner (1989) give a FORTRAN subroutine to calculate restricted range ($0 \leq \rho_{\max} \leq 1$) Tail and S/C intervals.

EXAMPLE: Kupper and Haseman (1978) report data from a teratology experiment in which ten pregnant mice were exposed to a certain chemical. The outcome of interest was the number of dead or malformed fetuses in each litter. Previous experience with such data show definite but low within-litter correlation. The data (y_i/n_i) are (0/5, 2/5, 1/7, 0/8, 2/8, 3/8, 0/9, 4/9, 1/10, 6/10) with $T = 10$, $y_+ = 19$, and $n_+ = 79$. The maximum likelihood estimate of (π, ρ) is (0.233, 0.148). Table 1 presents 90% confidence intervals for π computed by various methods.

Table 1 about here

The approximate beta-binomial confidence intervals described in Section 1 extend only slightly higher and lower than the small sample Tail and S/C intervals which assume independence ($\rho = 0$). They are suspected of being too short in this case. The unrestricted intervals are much longer and pulled toward $1/2$ than the remaining intervals, and would not be recommended here. However, the Tail or S/C intervals restricted to $\rho \leq 0.25$ might be appropriate. That both methods produce nearly identical intervals is suggested by the expected length plots in Figure 3. This figure shows that S/C intervals enjoy the greatest (expected) length advantage over the Tail intervals for π near $1/2$. The

TABLE 1

Various 90% interval estimates of π for the
teratological study reported by Kupper and Haseman (1978)

<u>Method</u>	<u>Confidence Interval</u>
Gladen	[.118, .363]
Southward–Van Ryzin	[.119, .343]
LR	[.134, .365]
Tail (unrestricted)	[.08, .51]
Tail ($\rho \leq 0.25$)	[.12, .40]
Tail ($\rho = 0$)	[.16, .33]
S/C (unrestricted)	[.11, .50]
S/C ($\rho \leq 0.25$)	[.13, .42]
S/C ($\rho = 0$)	[.17, .33]

teratology example has only a moderate π and small ρ which suggests the two methods will produce similar intervals. $\square$

The results of this paper suggest several areas for further research. First, the S/C algorithm above forces the addition or deletion of sets such as $\{\underline{y} \in \mathcal{Y} | y_+ = j\}$ to the acceptance sets. Another way to reduce the conservativeness of the S/C intervals is to allow the addition or deletion of subsets of $\{\underline{y} \in \mathcal{Y} | y_+ = j\}$ and similar modifications could be made to the Tail method (Santner and Snell, 1980; Wypij, 1987). Second, the Tail and S/C intervals as presented only depend on Y_+ and do not use all of the information contained in the sufficient statistic $(Y_1, \dots, Y_T)$. An alternate approach would be to construct a joint confidence region for (π, ρ) . The computations involved in such schemes may be enormous, however. Lastly, we note that this methodology can be extended to other "nuisance parameter" cases, including other correlated binomial or multinomial models and two group comparisons of proportions arising from correlated binary observations. For example, one could construct a confidence interval for the difference between the marginal probabilities of success of the teratology experiment having both the treatment and control groups of Kupper and Haseman (1978).

In summary, we suggest that the Tail and S/C intervals be used whenever T or n_+ is small, or for moderate values of T whenever π or ρ are extreme. In these cases the use of asymptotic intervals is suspect. The S/C intervals are less conservative and of shorter average length than the Tail intervals and so appear to be preferable, though they are more difficult to compute.

ACKNOWLEDGEMENTS

The work of the first author was partially supported by the Cornell University/Hospital for Special Surgery Program in Biomechanical Engineering and by NIEHS Training Grant ES-07142 at Harvard University. The work of the second author was partially supported by the U.S. Army Research Office through the Mathematical Sciences Institute of Cornell University.

REFERENCES

- Blyth, C.R. and Still, H.A. (1983). Binomial Confidence Intervals. J. Amer. Statist. Assoc. **76**, 108–116.
- Casella, G. (1986). Refining Binomial Confidence Intervals. Canad. J. Statist. **14**, 113–129.
- Chatfield, C. and Goodhardt, G.J. (1970). The Beta–Binomial Model for Consumer Purchasing Behavior. Appl. Statist. **19**, 240–250.
- Clopper, C.J. and Pearson, E.S. (1934). The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. Biometrika **26**, 404–413.
- Cox, D.R. and Hinkley, D. (1974). Theoretical Statistics. London: Chapman and Hall.
- Crow, E.L. (1956). Confidence Intervals for a Proportion. Biometrika **43**, 423–435.
- Crow, E.L. and Miles, M. (1979). Approximate Confidence Intervals for a Proportion from Markov Dependent Trials. Comm. Statist. B **8**, 25–52.
- Duffy, D.E. and Santner, T.J. (1987). Confidence Intervals for a Binomial Parameter Based on Multi–Stage Tests. Biometrics **43**, 81–93.
- Eaton, E.L. (1982). A Review of Selected Topics in Multivariate Probability Inequalities. Ann. Statist. **10**, 11–43.
- Gladden, B. (1979). The Use of the Jackknife to Estimate Proportions from Toxicological Data in the Presence of Litter Effects. J. Amer. Statist. Assoc. **74**, 278–283.
- Griffiths, D.A. (1973). Maximum Likelihood Estimation for the Beta–Binomial Distribution and an Application to the Household Distribution of the Total Number of Cases of a Disease. Biometrics **29**, 637–648.
- Kupper, L.L. and Haseman, J.K. (1978). The Use of a Correlated Binomial Model for the Analysis of Certain Toxicological Experiments. Biometrics **34**, 69–76.
- Marshall, A.W. and Olkin, I. (1979). Inequalities: Theory of Majorization and Its Applications. New York: Academic Press.
- Otake, M. and Prentice, R.L. (1984). The Analysis of Chromosomally Aberrant Cells Based on Beta–Binomial Distribution. Radiation Research **98**, 456–470.
- Prentice, R.L. (1986). Binary Regression Using an Extended Beta–Binomial Distribution, with Discussion of Correlation Induced by Covariate Measurement Errors. J. Amer. Statist. Assoc. **81**, 321–327.
- Santner, T.J. and Snell, M.K. (1980). Small–Sample Confidence Intervals for $p_1 - p_2$ and p_1/p_2 in 2×2 Contingency Tables. J. Amer. Statist. Assoc. **75**, 386–394.

- Skellam, J.G. (1948). A Probability Distribution Derived from the Binomial Distribution by Regarding the Probability of Success as Variable Between the Sets of Trials. J. Roy. Statist. Soc. B **10**, 257–261.
- Southward, G.M. and Van Ryzin, J. (1972). Estimating the Mean of a Random Binomial Parameter. Proc. Sixth Berkeley Symp. Math. Statist. Prob. **4**, 249–263.
- Sterne, T.E. (1954). Some Remarks on Confidence or Fiducial Limits. Biometrika **41**, 275–278.
- Van Ryzin, J. (1975). Estimating the Mean of a Random Binomial Parameter with Trial Size Random. Sankhyā Ser. B **37**, 10–27.
- Williams, D.A. (1975). The Analysis of Binary Responses from Toxicological Experiments Involving Reproduction and Teratogenicity. Biometrics **31**, 949–952.
- Wypij, D. (1987). Estimation Methods for Grouped Binary Data. Ph.D. Thesis, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York.
- Wypij, D. and Santner, T.J. (1989). Algorithms for Computing Interval Estimates of Beta–Binomial Success Probabilities. In preparation.

FIGURE LEGENDS

FIGURE 1

Achieved Coverage Probability of Gladen, Southward and Van Ryzin (both solid) and Likelihood Ratio (dashed) 90% Nominal Confidence Intervals as a Function of π for $T = 10$ Groups of Size n_i when $\rho = .5$

FIGURE 2

Achieved Coverage Probability of Tail (solid) and S/C (dashed) 90% Nominal Confidence Intervals as a Function of π for $T = 10$ Groups of Size $n_i = 3$ when $\rho = .5$

FIGURE 3

Expected Lengths of Tail (solid) and S/C (dashed) 90% Nominal Confidence Intervals as a Function of π for $T = 10$ Groups of Size $n_i = 3$ when $\rho = .5$

FIGURE 4

Achieved Coverage Probability of 90% Nominal Confidence Unrestricted ($\rho_{\max} = 1$) S/C Intervals (solid) and Restricted ($\rho_{\max} = .5$) S/C Intervals (dashed) as a Function of π for $T = 10$ Groups of Size $n_i = 3$ when $\rho = .5$

FIGURE 1

FIGURE 2

FIGURE 3

FIGURE 4

APPENDIX 1

Both parts of Theorem 2.1 hold, provided $P[Y_+ \geq j | \pi, \rho]$ is nondecreasing in π for every $\rho \in [0, 1]$. When $j = 0$ then $P[Y_+ \geq 0 | \pi, \rho] = 1$ for all $0 \leq \pi \leq 1$, $0 \leq \rho \leq 1$, and so the result holds trivially. When $j \geq 1$, and $\rho = 0$ then

$$P[Y_+ \geq j | \pi, 0] = \sum_{k=j}^n \binom{n}{k} \pi^k (1-\pi)^{n-k}$$

which is well known to be nondecreasing in π . A similar argument shows the result for $\rho = 1$. The proof of the result for $j \geq 1$ and $0 < \rho < 1$ is an application of the following known lemma. (See Theorems 3.1 and 3.2 of Eaton (1982).)

LEMMA A.1: Suppose $g_1(\cdot)$ and $g_2(\cdot)$ are probability densities with respect to measure μ on $\mathbb{R}^T$ and that

$$g_1(\underline{w})g_2(\underline{x}) \leq g_1(\underline{w} \wedge \underline{x})g_2(\underline{w} \vee \underline{x}) \quad (\text{A.1})$$

where $(\underline{w} \wedge \underline{x})_i = \min \{x_i, w_i\}$ and $(\underline{w} \vee \underline{x})_i = \max \{w_i, x_i\}$ for $1 \leq i \leq T$. Then

$$\int q(\underline{x})g_1(\underline{x})\mu(d\underline{x}) \leq \int q(\underline{x})g_2(\underline{x})\mu(d\underline{x})$$

for any nonnegative bounded function $q(\cdot)$ that is nondecreasing in each component.

If $T = 1$ and $g_2(x)/g_1(x)$ is nondecreasing in x then (A.1) holds. If $T > 1$ and $g_m(\underline{x}) = \prod_{i=1}^T g_m^*(x_i)$ for $m = 1, 2$ then (A.1) holds provided $g_2^*(w)/g_1^*(w)$ is nondecreasing in w .

Applying the law of conditional expectation gives

$$P[Y_+ \geq j | \pi, \rho] = \int q(p_1, \dots, p_T) \prod_{i=1}^T g(p_i | \pi, \rho) dp_i$$

where $g(\cdot | \pi, \rho)$ is the beta distribution with $\alpha = \pi(\rho^{-1} - 1)$ and $\beta = (1 - \pi)(\rho^{-1} - 1)$, and $q(\underline{p})$ is the conditional probability of $[Y_+ \geq j]$ when $Y_1, \dots, Y_T$ are mutually independent with Y_i distributed binomial with n_i trials and having success probability p_i . If $\pi_1 < \pi_2$ and $0 < \rho < 1$ then $g(p | \pi_2, \rho) / g(p | \pi_1, \rho) = C[p / (1 - p)]^{(\pi_2 - \pi_1)\rho / (1 - \rho)}$ where $C = C(\pi_1, \pi_2, \rho)$ is independent of p . The conclusion follows from Lemma A.1 provided $q(\underline{p})$ is nondecreasing in each p_i since $0 \leq q(\underline{p}) \leq 1$ for all $\underline{p} \in [0, 1]^T$. Without loss of generality it suffices to consider p_1 when all other p_i 's are fixed. Write

$$[\underline{y} \in \mathcal{Y} : y_+ \geq j] = A_0 \cup \left[\bigcup_{k=1}^{n_1} (A_k \cap [\underline{y} \in \mathcal{Y} : y_1 \geq k]) \right]$$

where $A_0 = [\underline{y} \in \mathcal{Y} : \sum_{i=2}^T y_i \geq j]$ and $A_k = [\underline{y} \in \mathcal{Y} : \sum_{i=2}^T y_i = j - k]$ for $1 \leq k \leq n_1$. Observe $\{A_k\}_{k=0}^{n_1}$ are pairwise disjoint. Therefore

$$P[Y_+ \geq j | \underline{p}] = P[A_0 | \underline{p}] + \sum_{k=1}^{n_1} P[A_k | \underline{p}] P[Y_1 \geq k | \underline{p}]$$

since Y_1 is independent $\sum_{i=2}^T Y_i$. However, $P[A_k | \underline{p}]$ is independent of p_1 for $0 \leq k \leq n_1$ and $P[Y_1 \geq k | \underline{p}]$ is strictly increasing in p_1 . These facts yield the result. $\square$

APPENDIX 2

LEMMA A.2: For any $0 \leq L \leq n_1/2$,

$$q(C[L, n_1 - L] | \pi) = \inf_{0 \leq \rho \leq 1} P[L \leq Y_1 \leq n_1 - L | \pi, \rho]$$

is nondecreasing in π on $[0, 1/2]$.

PROOF: It suffices to show for any ρ , $0 \leq \rho \leq 1$, and $0 \leq \pi_1 < \pi_2 \leq 1/2$ that

$$P[L \leq Y_1 \leq n_1 - L | \pi_1, \rho] \leq P[L \leq Y_1 \leq n_1 - L | \pi_2, \rho].$$

Regarding the probabilities as functions of the beta distribution parameters $\alpha = \pi(\rho^{-1} - 1)$ and $\beta = (1 - \pi)(\rho^{-1} - 1)$, this is equivalent to proving if (α_1, β_1) majorizes (α_2, β_2) (Marshall and Olkin (1979), p. 7, hereafter referred to as M/O) then

$$P[L \leq Y_1 \leq n_1 - L | \alpha_1, \beta_1] \leq P[L \leq Y_1 \leq n_1 - L | \alpha_2, \beta_2] \quad (\text{A.2})$$

since $\alpha_1 + \beta_1 = \beta_2 + \alpha_2$ and $\alpha_1 < \alpha_2$ implies the correlation is the same for the two distributions and the (α_1, β_1) distribution has smaller marginal success probability. By result E.20 of M/O the vector $(Y_1, n_1 - Y_1)$ forms a Schur family in (α, β) . The event $B = \{(k, n_1 - k) : k < L \text{ or } k > n - L\}$ satisfies property (3) of Proposition E.2 of M/O. Hence

$$P[Y_1 < L \text{ or } Y_1 > n_1 - L | \alpha_1, \beta_1] \geq P[Y_1 < L \text{ or } Y_1 > n_1 - L | \alpha_2, \beta_2]$$

which yields (A.2) and proves the lemma. $\square$

Intuitively, this result states that the probability of central outcomes increases as π approaches $1/2$ for any fixed value of the within-group correlation. It suggests the following more general result.

CONJECTURE: For any $T \geq 2$, $0 \leq L \leq n_+/2$,

$$q(C[L, n_+ - L] | \pi) = \inf_{0 \leq \rho \leq 1} P[L \leq Y_+ \leq n_+ - L | \pi, \rho]$$

is nondecreasing in π on $[0, 1/2]$.

The conjecture follows if for every $\rho \in [0, 1]$, $0 \leq \pi_1 < \pi_2 \leq 1/2$, and $0 \leq L \leq n_+/2$ it holds that

$$P[L \leq Y_+ \leq n_+ - L | \pi_1, \rho] \leq P[L \leq Y_+ \leq n_+ - L | \pi_2, \rho]. \quad (\text{A.3})$$

Several cases of (A.3) can be proved analytically. It is trivially true for $L = 0$. It is straightforward to prove (A.3) for $\rho = 0$ or 1 since Y_+ has (essentially) the binomial distribution in these cases. (A.3) can also be proven for the case $L = 1$ by checking the sign of the derivative of $P[Y_+ = 0 | \pi, \rho] + P[Y_+ = n_+ | \pi, \rho]$. Empirical calculation for other ρ , T , and L values support the result but its analytic proof is an open question.