Confidence Sets for Binary Response Models

ROBERT A. AGNEW*

Inequalities of Hoeffding are employed to obtain a confidence set for an arbitrary parameterized binary response model. Numerical results are demonstrated for a particular set of experimental data.

1. INTRODUCTION

Binary (or quantal) response refers to the experimental situation where the basic independent observations are numerically either zero or one, or descriptively either "failure" or "success." In a parameterized binary response model, the probabilities of success associated with these basic binary observations are linked via functional dependence on an unknown general parameter θ, assumed to lie in a parameter space T. A $100(1-\alpha)$-percent confidence set is then a statistical subset of T, depending on the basic binary observations, which covers the "true" parameter θ with probability not less than $1-\alpha$, no matter what the value of θ actually is. Alternatively, α is an upper bound on the Type I error associated with the hypothesis test: reject $\theta=\theta_0$ if and only if θ_0 is not covered by the confidence set.

A binary response model can generally be structured (parameterized) in various ways, and given a specific structure, alternative confidence sets can be defined. Often, there exists an "exact, optimum" (UMP) confidence set, but just about as often, this optimum set is computationally intractable. Our method of constructing a confidence set is inexact, nonoptimal, and perhaps rather crude. On the other hand, our confidence set is minimally complex, and we might add that our approach is motivated by the general scheme in Belyaev [2].

The statistical theory of binary response models has recently been summarized in Cox's monograph [4]. With primary emphasis on the "logistic" model, Cox reviews both point and set estimation methods. In the latter case, he details both exact and nominal (or asymptotically valid) confidence sets, but the exact sets are computationally cumbersome, even when the number of (real) parameters is only two.

In subsequent sections, we generalize the usual (two-sided, nonrandomized) confidence interval for the binomial probability of success to the "average probability of success" in the multi-parameter case via some inequalities of Hoeffding [6]. This generalization immediately yields a confidence set for an arbitrary parameterized binary response model. In Section 2 the usual binomial confidence interval is reviewed. Section 3 generalizes the usual interval to the multi-parameter case. Section 4 indicates the application to parameterized binary response models. In Section 5, the results of Section 4 are applied to a particular set of experimental data from Pickens [7].

We shall employ the following notation. For $n \geq 1$ and $0 \leq p \leq 1$, $S \sim b(n,p)$ implies that S is a binomially distributed random variable with parameters n and p, the probability of "success." $B(\cdot\,;n,p)$ denotes the corresponding distribution function; i.e., $P\{S \leq k\} = B(k;n,p)$ for $k=0,\cdots,n$. For $0<p<1$, $B(\cdot\,;n,p)$ is strictly increasing. On the other hand, $B(k;n,\cdot)$ is strictly decreasing for $k<n$. Moreover, $B(k-1;n,k/n) \leq \frac{1}{2} \leq B(k;n,k/n)$ for $k=1,\cdots,n$ by Corollaries 2.1, 2.2 (see [1, p. 2–3]) plus the Central Limit Theorem. For any nondecreasing function f mapping the integers $\{0,\cdots,n\}$ into the unit interval $[0,1]$, let $f^{-1}(p) = \max\{k : f(k) \leq p\}$ for $f(0) \leq p \leq f(n)$. Then, $f(f^{-1}(p)) \leq p < f(f^{-1}(p)+1)$ for $f(0) \leq p < f(n)$.

2. THE BINOMIAL CONFIDENCE INTERVAL

Fix $n \geq 1$, and let $0<\alpha<\frac{1}{2}$. Define the increasing functions f_1 and f_2 mapping $\{0,\cdots,n\}$ into $[0,1]$ by: $f_1(0)=0$; $B(k-1;n,f_1(k)) = 1-\alpha$ for $k=1,\cdots,n$; $B(k;n,f_2(k)) = \alpha$ for $k=0,\cdots,n-1$; and $f_2(n)=1$. Then $f_1(k) < k/n$ for $k=1,\cdots,n$ and $f_2(k) > k/n$ for $k=0,\cdots,n-1$. Moreover, $S \sim b(n,p)$ implies

$$P\{f_1(S) \leq p\} = B(f_1^{-1}(p);n,p) \geq B(f_1^{-1}(p);n,f_1(f_1^{-1}(p)+1)) = 1-\alpha$$

for $0<p<f_1(n)$ and

$$P\{f_2(S) \leq p\} = B(f_2^{-1}(p);n,p) \leq B(f_2^{-1}(p);n,f_2(f_2^{-1}(p))) = \alpha$$

for $f_2(0)<p<1$. Hence,

$$P\{f_1(S) \leq p \leq f_2(S)\} \geq 1-2\alpha,$$

so that $[f_1(S), f_2(S)]$ defines a $100(1-2\alpha)$-percent confidence interval for p.

Many statistics books (e.g., Zehna [9, p. 377–80]) indicate how $f_1(k)$ and $f_2(k)$ can be obtained (for a given k and standard α value) from beta distribution tables. Making use of the identity

$$B(k-1;n,p) = 1 - I(p;k,n-k+1) = I(1-p;n-k+1,k)$$

* Robert A. Agnew is operations research project manager, Corporate Research Department, Montgomery Ward and Co., Chicago, Ill. 60680. The author was formerly assistant professor, Department of Systems Management, Air Force Institute of Technology.

© Journal of the American Statistical Association
June 1974, Volume 69, Number 346
Theory and Methods Section

for $k = 1, \cdots, n$, where

$$I(p; r, s) = ((r + s - 1)!/(r - 1)!(s - 1)!) \cdot \int_0^p u^{r-1}(1 - u)^{s-1} du$$

for positive integers r and s and $0 \leq p \leq 1$, we have:

$$I(f_1(k); k, n - k + 1) = \alpha$$

for $k = 1, \cdots, n$;

$$I(1 - f_2(k); n - k, k + 1) = \alpha$$

for $k = 0, \cdots, n - 1$; and the desired values are available in [3, p. 251–65] or more precisely in [5].

Example

Suppose that $n = 23$, $S = 15$, and $\alpha = .05$. From Beyer [3, p. 259], with interpolation, we have

$$f_1(15) \cong (.4)(.49816) + (.6)(.43711)$$
$$1 - f_2(15) \cong (.8)(.19556) + (.2)(.15682)$$

or $f_1(15) \cong .46$ and $f_2(15) \cong .81$. Precise values from [5] are $f_1(15) = .4595441$ and $f_2(15) = .8136562$.

3. THE GENERALIZED CONFIDENCE INTERVAL

Now suppose that $S_1, \cdots, S_m$ are independent random variables with $S_i \sim b(n_i, p_i)$. Let

$$n = \sum_{i=1}^m n_i \quad \text{and} \quad S = \sum_{i=1}^m S_i.$$

We want to obtain a confidence interval for

$$p = n^{-1} \sum_{i=1}^m n_i p_i,$$

the "average probability of success."

Hoeffding [6, Theorem 4, p. 718] proved that

$$\begin{aligned} P\{S \leq k\} &\geq B(k; n, p) \quad (k \geq np), \\ P\{S \leq k\} &\leq B(k; n, p) \quad (k \leq np - 1) \end{aligned} \tag{3.1}$$

(see also [1, p. 2]).

Hence, the confidence interval from Section 2 would carry over if we had $f_1^{-1}(p) \geq np$ for $0 < p < f_1(n)$ and $f_2^{-1}(p) \leq np - 1$ for $f_2(0) < p < 1$. However, this is not the case; e.g., $f_1^{-1}(p) = 0$ for p sufficiently small. It is necessary to correct for deficiencies associated with p values near zero and one.

Define the nondecreasing functions g_1 and g_2 mapping $\{0, \cdots, n\}$ into $[0, 1]$ by: $g_1(0) = 0$;

$$g_1(k) = \min (f_1(k), (k - 1)/n)$$

for $k = 1, \cdots, n$;

$$g_2(k) = \max (f_2(k), (k + 1)/n)$$

for $k = 0, \cdots, n - 1$; and $g_2(n) = 1$. We have then that $g_1^{-1}(p) \geq \max (f_1^{-1}(p), np)$ for $0 < p < g_1(n)$ and $g_2^{-1}(p) \leq \min (f_2^{-1}(p), np - 1)$ for $g_2(0) < p < 1$. Hence, it follows from the construction of f_1 and f_2 and Hoeffding's inequalities [6, (3.1)] that

$$P\{g_1(S) \leq p\} \geq B(g_1^{-1}(p); n, p) \geq B(f_1^{-1}(p); n, p) \geq 1 - \alpha$$

for $0 < p < g_1(n)$ and

$$P\{g_2(S) \leq p\} \leq B(g_2^{-1}(p); n, p) \leq B(f_2^{-1}(p); n, p) \leq \alpha$$

for $g_2(0) < p < 1$. We conclude that $[g_1(S), g_2(S)]$ defines a $100(1 - 2\alpha)$-percent confidence interval for p. Generally, this interval will coincide with the usual one except for extreme S values near 0 or n; e.g., $g_1(1) = 0$ and $g_2(n - 1) = 1$.

Example

Suppose again that $n = 23$, $S = 15$, and $\alpha = .05$. We have

$$g_1(15) = \min (f_1(15), 14/23) \cong \min (.46, .61) = .46$$
$$g_2(15) = \max (f_2(15), 16/23) \cong \max (.81, .70) = .81$$

so the two 90-percent confidence intervals coincide.

4. APPLICATION TO BINARY RESPONSE MODELS

In a parameterized binary response model, we have the setup of Section 3 with $p_i = \lambda_i(\theta)$, where θ is an unknown general parameter, assumed to lie in a parameter space T, and the λ_i are known functions mapping T into the unit interval. Assuming n_i independent binary replicates for case i ($i = 1, \cdots, m$) and total independent binary responses $n = \sum_{i=1}^m n_i$, we form the "average probability of success function"

$$\lambda = n^{-1} \sum_{i=1}^m n_i \lambda_i \tag{4.1}$$

and the confidence set

$$H(S) = \{t \in T: g_1(S) \leq \lambda(t) \leq g_2(S)\} \tag{4.2}$$

where S is the total number of successes in n binary trials. From the development in Section 3, it is clear that $P\{\theta \in H(S)\} \geq 1 - 2\alpha$, so that $H(S)$ defines a $100(1 - 2\alpha)$-percent confidence set for θ.

The most common experimental situation involves the estimation of confidence bounds for a binary (or quantal) response "curve." In this case, we have $p_i = F^*(x_i)$, where $F^*(x)$ is the probability of success associated with known "dosage" or "score" $x \in R$, the real numbers. The function F^* is unknown, but is assumed *a priori* to be an element of $\mathfrak{F}$, the set of all nondecreasing, right-continuous functions mapping R into the unit interval. The problem is parameterized by assuming $F^*(\cdot) = G(\cdot; \theta)$, where $\{G(\cdot; t): t \in T\} \subset \mathfrak{F}$, and $\lambda_i(t) = G(x_i; t)$.

If we put

$$\begin{aligned} \bar{F}(x; S) &= \sup \{G(x; t): t \in H(S)\} \\ \underline{F}(x; S) &= \inf \{G(x; t): t \in H(S)\}, \end{aligned} \tag{4.3}$$

then clearly

$$\bigcap_{x \in R} \{\underline{F}(x; S) \leq F^*(x) \leq \bar{F}(x; S)\} \supset \{\theta \in H(S)\}$$

as events so that

$$H'(S) = \{F \in \mathfrak{F} : \underline{F}(\cdot\,; S) \leq F(\cdot) \leq \bar{F}(\cdot\,; S)\} \quad (4.4)$$

defines a derivative $100(1-2\alpha)$-percent confidence set (band, region) for the "true" function F^*. This confidence set can be "graphed" as the band between the two monotone functions $\underline{F}(\cdot\,; S)$ and $\bar{F}(\cdot\,; S)$.

It is commonly assumed that $G(x; t_1, t_2) = \Phi(t_1 + t_2 x)$ with $T = \{(t_1, t_2) : t_2 \geq 0\}$, where Φ is some standardized cumulative distribution function like the logistic or normal. Although such cases can be treated by Lagrangian type methods, a more general parameterization simply puts $T = \mathfrak{F}$, $G(x; t) = t(x)$, and $\theta = F^*$.

With this general parameterization, denote

$$y = n^{-1} \sum_{i=1}^{m} n_i y_i$$

where $y_i = t(x_i) = \lambda_i(t)$ and assume $x_1 \leq \cdots \leq x_m$. Then,

$$\bar{F}(x; S) = \min_i \max \{y_i : 0 \leq y_1 \leq \cdots \leq y_m \leq 1; \; g_1(S) \leq y \leq g_2(S); x \leq x_i\}$$

$$= \min (1, \min_i \{(n g_2(S)/\textstyle\sum_{j=i}^{m} n_j); x \leq x_i\}) \quad (4.5)$$

for $x \leq x_m$, and $\bar{F}(x; S) = 1$ for $x > x_m$. Similarly,

$$\underline{F}(x; S) = \max_i \min \{y_i : 0 \leq y_1 \leq \cdots \leq y_m \leq 1; \; g_1(S) \leq y \leq g_2(S); x \geq x_i\}$$

$$= \max (0, \max_i \{1 - (n(1 - g_1(S))/\textstyle\sum_{j=1}^{i} n_j) : x \geq x_i\}) \quad (4.6)$$

for $x \geq x_1$, and $\underline{F}(x; S) = 0$ for $x < x_1$.

5. A NUMERICAL EXAMPLE

Consider the following data obtained from Pickens [7]. The x_i are Quantitative Aptitude/Graduate Record Examination scores for students in a particular graduate engineering program and "success" means attainment of a Master of Science degree.

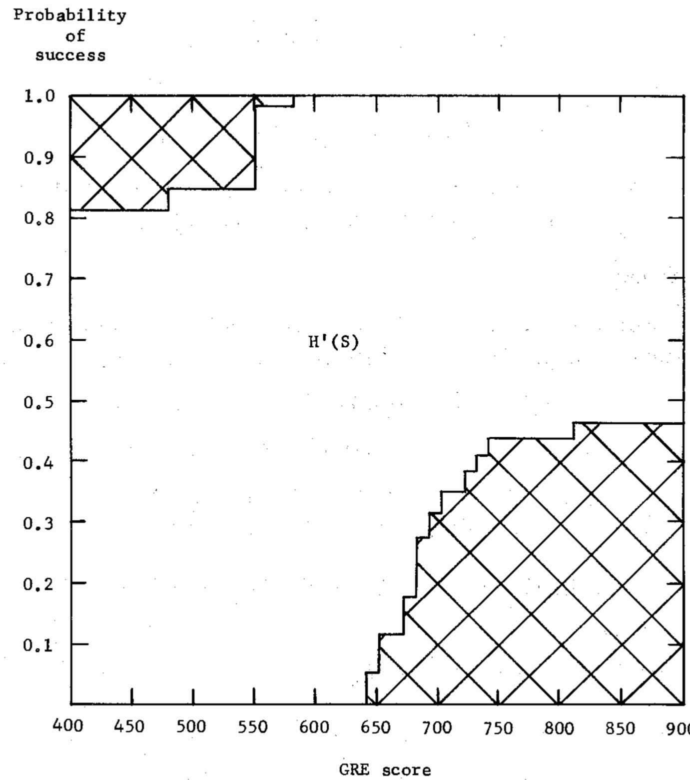
We have $n = 23$, $S = 15$, so that $g_1(S) = f_1(S) \cong .46$ and $g_2(S) = f_2(S) \cong .81$ from our previous examples.

Experimental Data

i	x_i	n_i	S_i
1	480	1	0
2	550	3	0
3	580	3	2
4	590	2	1
5	610	2	2
6	630	1	1
7	640	1	0
8	650	1	1
9	670	1	1
10	680	2	2
11	690	1	0
12	700	1	1
13	720	1	1
14	730	1	1
15	740	1	1
16	810	1	1

Employing the general parameterization of Section 4, we can display our 90-percent confidence set for the true probability of success curve as the non-crosshatched area in the following figure.

Confidence Set



[*Received March 1972. Revised November 1973.*]

REFERENCES

[1] Anderson, T.W. and Samuels, S.M., "Some Inequalities Among Binomial and Poisson Probabilities," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, Berkeley: University of California Press, 1967, 1–12.

[2] Belyaev, Y.K., "On Confidence Intervals and Sets for Various Statistical Models," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, Berkeley: University of California Press, 1967, 51–8.

[3] Beyer, W.H. (ed.), *Handbook of Tables for Probability and Statistics*, 2nd ed., Cleveland: The Chemical Rubber Co., 1968.

[4] Cox, D.R., *The Analysis of Binary Data*, London: Methuen and Co., Ltd., 1970.

[5] Harter, H.L., *New Tables of the Incomplete Gamma-Function Ratio and of Percentage Points of the Chi-Square and Beta Distributions*, Aerospace Research Laboratories, U.S. Government Printing Office, 1964.

[6] Hoeffding, W., "On the Distribution of the Number of Successes in Independent Trials," *Annals of Mathematical Statistics*, 27 (September 1956), 713–21.

[7] Pickens, T.A., "A Logistic Quantal Response Model for Estimating the Probability of Graduate Success," Unpublished M.S. thesis, Department of Systems Management, Air Force Institute of Technology, 1971.

[8] Scheffé, H., *The Analysis of Variance*, New York: John Wiley and Sons, Inc., 1959.

[9] Zehna, P.W., *Probability Distributions and Statistics*, Boston: Allyn and Bacon, Inc., 1970.