Journal of Modern Applied Statistical Methods

Volume 18 | Issue 1 Article 29

6-1-2020

Inferences About the Probability of Success, Given the Value of a Covariate, Using a Nonparametric Smoother

Rand Wilcox
University of Southern California, rwilcox@usc.edu

Journal of Modern Applied Statistical Methods
May 2019, Vol. 18, No. 1, eP3252
doi: 10.22237/jmasm/1556670240

בס"ד

ISSN 1538 − 9472

INVITED ARTICLE

Inferences About the Probability of Success, Given the Value of a Covariate, Using a Nonparametric Smoother

Rand Wilcox
University of Southern California
Los Angeles, CA

For a binary random variable Y, let $p(x) = P(Y = 1 \mid X = x)$ for some covariate X. The goal of computing a confidence interval for $p(x)$ is considered. In the logistic regression model, even a slight departure difficult to detect via a goodness-of-fit test can yield inaccurate results. The accuracy of a confidence interval can deteriorate as the sample size increases. The goal is to suggest an alternative approach based on a smoother, which provides a more flexible approximation of $p(x)$.

Keywords: binary data, categorical data, logistic regression, Agresti-Coull method, Clopper-Pearson

Introduction

Consider the random variables X and Y having some unknown bivariate distribution, where $Y = 0$ or 1. As is evident, a fundamental goal is computing a $1 - \alpha$ confidence interval for $p(x) = P(Y = 1 \mid X = x)$. One approach (e.g., Piegorsch & Casella, 1988) is based on the logistic regression model. Assume for some unknown parameters β_0 and β_1,

$$p\left(x\right) = \frac{\exp\left(\beta_0 + \beta_1 x\right)}{1 + \exp\left(\beta_0 + \beta_1 x\right)}. \tag{1}$$

Briefly, inferences about $p(x)$ are made by focusing on the logit transform $\phi(x) = \log\{p(x)/(1 - p(x)\}$ and then computing a confidence interval for $\phi(x)$. Then,

doi: 10.22237/jmasm/1556670240 | Accepted: June 14, 2019; Published: June 1, 2020.
Correspondence: Rand Wilcox, rwilcox@usc.edu

the reverse transform, $\exp(\phi(x)/(1 + \exp(\phi(x))$, can be applied to obtain a confidence interval for $p(x)$. This will be labeled method LT henceforth. For details, including how to compute a Scheffé-type confidence band, see Brand et al. (1973) and Khorasani and Milliken (1982).

An issue is whether (1) provides a sufficiently good fit given the goal of computing a confidence interval for $p(x)$. One strategy is to perform the goodness-of-fit test in Hosmer and Lemeshow (1989) or the method recommended by Hosmer et al. (1997), and if it fails to reject, compute a confidence interval as just indicated. However, there are at least two issues that must be addressed. First, do these goodness-of-fit tests have sufficient power to detect situations where the fit is inadequate given the goal of computing a confidence interval for $p(x)$? Second, if it is decided that the logistic regression model provides an unsatisfactory fit, what method might be used instead?

To provide some perspective on the first issue, consider the situation where $p(x)$ is given by the regression line in the left panel of Figure 1 and the goal is to compute a $1 - \alpha = 0.95$ confidence interval for $p(x)$. Here, X is taken to have a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 4$. Three choices for x are considered, namely estimates of the quartiles which are labeled $\hat{x}_{0.25}$, $\hat{x}_{0.5}$ and $\hat{x}_{0.75}$. A simulation based on 5000 replications was used to estimate the actual value of α when using method LT. (The computations were performed by the R function `glm`.) This was done for three sample sizes: 100, 200 and 400. Then these simulations were repeated, only now the regression line for $p(x)$ corresponds to the regression line in the right panel of Figure 1. Table 1 shows the results.

Table 1. Estimates of α when computing a $1 - \alpha = 0.95$ confidence interval using method LT. The columns refer to the regression line in the Left and Right panels, respectively, of Figure 1.

	Left			Right		
n	$\hat{x}_{0.25}$	$\hat{x}_{0.5}$	$\hat{x}_{0.75}$	$\hat{x}_{0.25}$	$\hat{x}_{0.5}$	$\hat{x}_{0.75}$
100	0.028	0.040	0.046	0.024	0.120	0.082
200	0.052	0.042	0.042	0.044	0.188	0.142
400	0.058	0.056	0.048	0.046	0.286	0.248

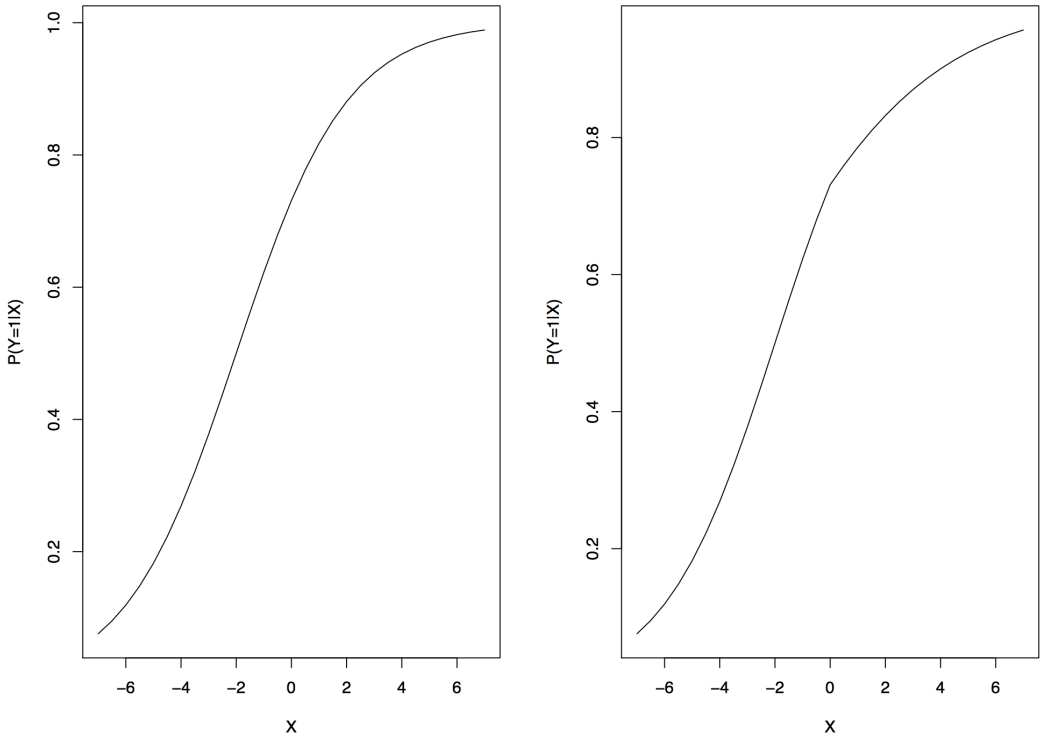

Figure 1. Regression lines used to illustrate the impact of a slight departure from the logistic regression model

For the left panel of Figure 1 the actual probability coverage is estimated to be close to the nominal level. This was expected because data were generated according to the logistic regression model given by (1) with $\beta_0 = 1$ and $\beta_1 = 0.5$. However, for the right panel, the accuracy of the confidence intervals deteriorates as the sample size increases and is highly unsatisfactory. The reason is the data were generated via (1) with $\beta_0 = 1$ and $\beta_1 = 0.5$ when $X < 0$, and $\beta_0 = 1$ and $\beta_1 = 0.3$ when $X \geq 0$. That is, the same regression line does not apply over the entire range of the explanatory variable. Hastie and Tibshirani (1990), Wilcox (2017), and others indicated specific values for the parameters of a linear model might suffice over some interval of the explanatory variable, but otherwise this is not the case. Examples based on data from two different studies are given below under Illustrations. What can be needed is a more flexible approach regarding how a regression line is fitted to the data.

Consider the strategy of trying to justify the logistic regression model based on a goodness-of-fit test. For the situation in the right panel of Figure 1, the method

in Hosmer and Lemeshow (1989) test was applied for the same situations used in Table 1. The computations were performed via the R function `hoslem.test` available in the R package `ResourceSelection`. For $n = 100$ and when testing at the 0.05, 0.10 and 0.20 levels, power was estimated to be 0.083, 0.147 and 0.261, respectively. That is, even testing at the 0.20 level, it is likely that this test will not detect the departure from the logistic regression model. For $n = 200$, now the estimates are 0.10, 0.168 and 0.295. For $n = 400$ the estimates are 0.152, 0.233 and 0.362. Using instead the goodness of fit test recommended by Hosmer et al. (1997), via the R package `rms`, for $n = 100$ power is 0.0986 0.146 and 0.242, again testing at the 0.05, 0.10 and 0.20 levels, respectively. For $n = 200$ the estimates are 0.131, 0.200 and 0.300; and for $n = 400$ the estimates are 0.199, 0.281 and 0.399. Power is improved using the method recommended by Hosmer et al. (1997), but again the likelihood of not detecting this departure from the logistic regression model is fairly high. Hence, the goal is to suggest an alternative method aimed at dealing with the limitation of the logistic regression model.

Preliminary Considerations

The strategy for computing a confidence interval is to search for a method that performs reasonably well in simulations without making any parametric assumptions about the nature of the association. The initial approach was to focus on the smoother in Wilcox (2017, section 11.5.8), which is a slight modification of the smoother in Hosmer and Lemeshow (1989).

For the random sample $(X_1,Y_1), \ldots, (X_n,Y_n)$, let $Z_i = (X_n - M) / \text{MADN}$, where M is the sample median based on $X_1, \ldots, X_n$, MAD is the median of $|X_1 - \text{M}|, \ldots, |\text{Xn} - \text{M}|$ and MADN = MAD / 0.6745 (under normality, MADN estimates the standard deviation). The estimate of $p(x)$ is taken to be

$$\hat{P}_{HL}\left(x\right) = \frac{\sum w_i Y_i}{\sum w_i}, \tag{2}$$

where

$$w_i = I_h e^{-\left(Z_i - z\right)^2},$$

the indicator function $I_h = 1$ if $|Z_i - z| < h$, otherwise $I_h = 0$, and $z = (x - M) / \text{MADN}$; h is called the span. The choice $h = 1.2$ appears to work well

in general (e.g., Wilcox, 2017) but clearly there are situations where some other choice is better (cf. Copas, 1983; Kay & Little, 1987). An expression for the standard error of $\hat{P}_{HL}(x)$ is easily derived (e.g., Fowlkes, 1987), which is a function of $P(Y = 1 \mid X = X_i)$ $(i = 1, \ldots, n)$.

Suppose the standard error of $\hat{P}_{HL}(x)$, τ, is known to a high degree of accuracy. An obvious strategy is to assume that $\hat{P}_{HL}(x)$ has, approximately, a normal distribution, in which case a $1 - \alpha$ confidence interval is taken to be $\hat{P}_{HL}(x) \pm c\tau$, where c is the $1 - \alpha/2$ quantile of a standard normal distribution. A natural speculation is a fairly large sample size might be needed so the actual probability coverage is reasonably close to the nominal level.

As a partial check, a simulation based on 10,000 replications was used to determine the standard error of $\hat{P}_{HL}(x)$ for the situation depicted in the left panel of Figure 1 when $n = 100$ and when x is taken to be the 0.75 quantile of the distribution of X. Then, a simulation was performed to estimate the actual probability coverage when computing a 0.95 confidence interval, which was estimated to be 0.86. Increasing n to 200, the estimate was 0.92. And of course there is the practical problem of obtaining a reasonably accurate estimate of the standard error. Consequently, this approach was abandoned.

Another approach, which has practical value in a range of similar situations (e.g., Wilcox, 2017), is to use a percentile bootstrap method. Briefly, generate a bootstrap sample by sampling with replacement n points from $(X_1,Y_1), \ldots, (X_n,Y_n)$. Based on this bootstrap sample compute $\hat{P}_{HL}(x)$. Repeat this process B times, put the estimates in ascending order, and use the middle $(1 - \alpha)B$ values to determine a $1 - \alpha$ confidence interval. Preliminary simulations indicated this approach works well provided x is not too far from the median. If x is taken to be the estimate of the 0.75 quantile, a sample size greater than 200 can be needed. For this reason, more precise details are not provided and this approach is not considered henceforth.

The Proposed Method

Consider the situation where Y is continuous and the goal is to estimate some robust measure of location associated with Y given that $X = x$, where $x \in \mathcal{I}$ and $\mathcal{I} = \{x: a < x < b\}$ for some specified constants a and b. When using a trimmed mean, numerous results indicate that a reasonably accurate confidence interval can be obtained by focusing on the points (X_i,Y_i) such that $a \leq X_i \leq b$. The Tukey and

McLaughlin (1963) method or a percentile bootstrap method may be used when a and b are determined based on the so-called running interval smoother (Wilcox, 2017). In essence, when there is an association, a reasonably accurate confidence interval can be computed provided that the length of the interval (a,b) is not too large This suggests that a similar approach might have practical value for the situation at hand.

The running interval smoother is essentially the smoother given by (2) but with a different choice for the weights and the span. The ith weight is $w_i = 1$ if $|Z_i - z| < h$; otherwise $w_i = 0$. Let $X = \{X_i: |Z_i - z| < h\}$ and $W = \Sigma_{X_i \in X} Y_i$. The estimate of $p(x)$ is $\hat{p}_S = W/m$, where m is the cardinality of the set X. That is, $\hat{p}_S$ is simply the proportion of successes given that $|Z_i - z| < h$. This will be called method S henceforth. If there is no association between Y and X, a confidence interval for $p(x)$ can be computed using extant methods based on a binomial distribution. However, choosing the span h so as to get a reasonable approximation of $p(x)$, and once the span has been chosen, it is necessary to find a method to provide a reasonably accurate confidence interval.

Consider the choice for the span, h. For extant smoothers (e.g., Hosmer & Lemeshow, 1989; Fowlkes 1987), there is no known computational method, based on the available data, for determining the span in a completely satisfactory manner. The best that can be done is to choose a value that appears to perform tolerably well for a reasonably broad range of situations and perhaps consider some additional values based on their impact on the plot of the regression line. Obviously this judgmental process is difficult to simulate.

The choice for h depends on the strength of the association between X and Y. When there is a relatively strong association, a small choice for h might be needed to get a good approximation of $p(x)$ at the expense of wider confidence intervals compared to using a relatively large h. Using $h = 1.2$, as done by the Hosmer–Lemeshow smoother, was found to be adequate when there is a very weak association. However, even for a moderately strong association, a much smaller value for h can be required.

Here, $h = 0.5$ is used unless stated otherwise. The idea is to use a value for h that provides a reasonable approximation of $p(x)$ even when there is a fairly strong association, with the understanding that for a very strong association, a smaller value for h might be preferable. To provide at least some perspective on this choice for h, consider the situation where the logistic regression model given by (1) is true with $\beta_1 = 0$. The value of $p(x)$ was computed for each of the deciles of X based on a sample size of 10,000. (As in Figure 1, X has a normal distribution with mean

zero and standard deviation 4.) The largest value for $|\hat{p}_S(x) - p(x)|$, when $h = 0.5$, was 0.016 and occurred when x is equal to the 0.1 quantile. When $\beta_1 = 0.5$, now the largest value for $|\hat{p}_S(x) - p(x)|$ is 0.03 and again occurred when x is equal to the 0.1 quantile. Lowering h to 0.45, the largest absolute difference is 0.025 and for $h = 0.4$ it is 0.013.

There is an extensive literature dealing with the goal of computing a confidence interval for the probability of success when dealing with a binomial distribution (e.g., Blyth, 1986; Brown et al., 2002; Schilling & Doi, 2014). A classic approach is the Clopper and Pearson (1934) method. If Y is assumed to have a binomial distribution when attention is restricted to those $X_i \in \mathcal{X}$, the Clopper-Pearson lower and upper ends of the confidence interval are $B(\alpha/2; W, m - W + 1)$ and $B(1 - \alpha/2; W + 1, m - W)$, respectively, where $B(q;u,v)$ is the qth quantile of a beta distribution with shape parameters u and v. The results reported by Blyth (1986) suggest using a method derived by Pratt (1968). Results in Brown et al. (2002) point to a method derived by Agresti and Coull (1998). The Agresti-Coull (AC) method is as follows. Let $\tilde{m} = m + z^2$ and $\tilde{W} = W + z^2/2$, where z is the $1 - \alpha/2$ quantile of a standard normal distribution. Let $p(x) = \tilde{W} / \tilde{m}$. The AC $1 - \alpha$ confidence interval for $p(x)$

$$\tilde{p}\left(x\right) \pm z\sqrt{\frac{\tilde{p}\left(x\right)\left(1-\tilde{p}\left(x\right)\right)}{\tilde{m}}} \tag{3}$$

Details about Pratt's method are not provided, because it did not perform well in simulations for the situations considered below. Even with a moderately large sample size the Schilling and Doi method has an extremely high execution time and is not considered further.

Preliminary simulations suggested CP performs well for $n \geq 80$, but for $n = 50$ it is too conservative; the actual probability coverage can be substantially higher than the nominal level. As for method AC, it was found to be unsatisfactory when $n \geq 100$ and indeed its performance was found to deteriorate as the sample size increases, in contrast to method CP. However, for $n = 50$, AC did perform well. When $n = 80$, methods CP and AC were found to perform about equally well. Consequently, it is assumed that AC is used when $n < 80$; when $n \geq 80$, CP is used. This will be called method SACCP henceforth.

Simulation Results

Simulations were used to study the finite sample properties of method SACCP for five situations, four of which are depicted in Figure 2. As in Figure 1, X is taken to have a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 4$. The basic strategy was to first generate data in a manner related to the logistic regression model based on a sample size of 20,000 and then compute $\hat{p}_S$ for five quantiles: 0.1, 0.25, 0.5, 0.75 and 0.9, which are –5.13, –2.70, 0.00, 2.67 and 5.13, respectively. These values were taken to be the true values of $p_S(x)$. Then a simulation with 5000 replications was used to determine how well method SACCP performs when computing a 0.95 confidence interval based on sample sizes $n = 50$, 100, and 200. The regression line related to the logistic regression model corresponds to the solid lines in Figure 2. The dotted lines correspond to $p_S(x)$ based on $n = 20{,}000$. For the first situation, S1, shown in the upper left panel of Figure 2, $p(x)$ corresponds to (1) with $\beta_0 = 1$ and $\beta_1 = 0.3$. In the upper right panel, $\beta_0 = 1$ and $\beta_1 = 0.5$, which is designated situation S2. The lower left panel (situation S3) is the same situation as depicted in the right panel of Figure 1. The lower right panel (S4) corresponds to

$$p(x) = \frac{\exp\left(1 + 0.5x + 0.1x^2\right)}{1 + \exp\left(1 + 0.5x + 0.1x^2\right)}. \tag{4}$$

The first issue here is whether accurate confidence intervals can be computed for the regression line $p_S(x)$, which is designed to provide a reasonable approximation of the true regression line $p(x)$ when the logistic regression model is incorrect. For the situation in the lower left panel of Figure 2, method LT performs poorly and the accuracy of confidence intervals deteriorates as n increases. The goal here is to provide some indication of whether accurate confidence intervals can be computed for the dotted lines in Figure 2. Some results related to confidence intervals for $p(x)$, when the logistic regression model is correct, are reported.

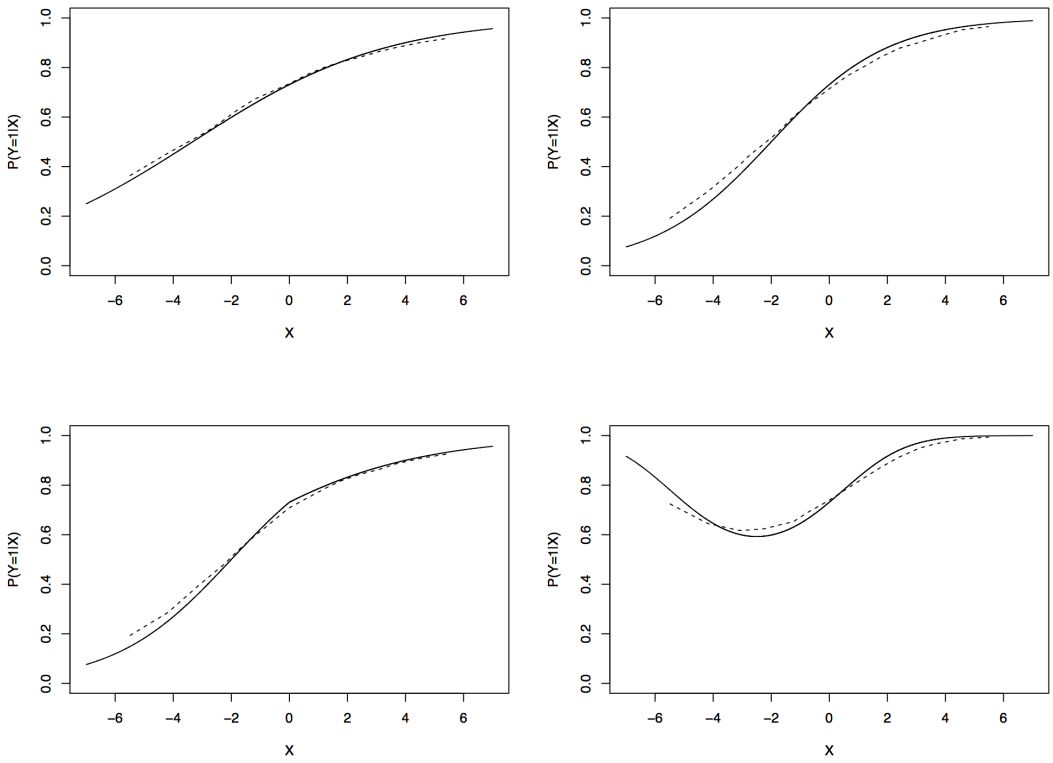

Figure 2. The solid line in the upper left panel is the logistic regression line when $\beta_0 = 1$ and $\beta_1 = 0.3$ (situation S1). The dotted line is the regression line based on method S. The upper right panel is when $\beta_0 = 1$ and $\beta_1 = 0.5$ (S2). The lower left is when $\beta_1 = 0.5$ for $X < 0$ and $\beta_1 = 0.3$ when $X > 0$ (S3). In the lower right panel (S4), $p(x)$ is given by equation (4).

Experience with smoothers (e.g., Wilcox, 2017) indicates that situations occur where there is a clear association between X and Y over some range of X values, but outside this range, the association appears to be substantially weaker with the possibility there is little or no association. Consequently, for the fifth situation (S5), data are generated according to (1) with $\beta_0 = 1$ and $\beta_1 = 0.3$ when $X < 0$. For $X \geq 0$, $\beta_1 = 0$ was used.

The results are reported in Table 2. The estimate of α never exceeds 0.05. The main difficulty is that some estimates drop below 0.025, particularly when dealing with the 0.9 quantile. For S4, the estimate is 0.001 with $n = 200$. Decreasing the span to 0.4, the estimate is 0.034.

Table 2. Estimates of α when using method SACCP to compute $1 - \alpha = 0.95$ confidence interval for the regression line estimated by method S.

Method	n	$x = -5.13$	$x = -2.70$	$x = 0$	$x = 2.70$	$x = 5.13$
S1	50	0.023	0.028	0.028	0.025	0.017
	100	0.023	0.036	0.003	0.030	0.029
	200	0.040	0.042	0.033	0.035	0.034
S2	50	0.017	0.023	0.024	0.025	0.017
	100	0.025	0.030	0.036	0.030	0.029
	200	0.033	0.035	0.034	0.030	0.034
S3	50	0.017	0.030	0.025	0.000	0.000
	100	0.025	0.030	0.035	0.027	0.013
	200	0.033	0.035	0.039	0.032	0.015
S4	50	0.019	0.028	0.027	0.009	0.003
	100	0.028	0.030	0.033	0.012	0.013
	200	0.035	0.042	0.040	0.026	0.001
S5	50	0.023	0.028	0.028	0.025	0.017
	100	0.032	0.036	0.037	0.030	0.029
	200	0.040	0.042	0.033	0.035	0.034

Table 3. Estimates of α when using method SACCP to compute $1 - \alpha = 0.95$ confidence interval for the regression line estimated by method SACCP when the logistic regression model is correct.

	n	$x = -5.13$	$x = -2.70$	$x = 0$	$x = 2.70$	$x = 5.13$
$\beta_1 = 0.1$, $h = 0.5$	50	0.019	0.024	0.024	0.023	0.012
	100	0.033	0.026	0.036	0.029	0.020
	200	0.030	0.039	0.036	0.035	0.029
	400	0.043	0.042	0.044	0.037	0.033
$\beta_1 = 0.1$, $h = 0.75$	50	0.028	0.029	0.030	0.026	0.018
	100	0.033	0.032	0.037	0.036	0.031
	200	0.040	0.046	0.038	0.043	0.041
	400	0.058	0.052	0.043	0.049	0.048
$\beta_1 = 0.3$, $h = 0.5$	50	0.020	0.027	0.024	0.017	0.017
	100	0.038	0.037	0.038	0.033	0.021
	200	0.052	0.045	0.036	0.040	0.032
	400	0.049	0.039	0.044	0.042	0.034
$\beta_1 = 0.5$, $h = 0.5$	50	0.032	0.030	0.030	0.021	0.016
	100	0.051	0.044	0.040	0.037	0.024
	200	0.086	0.060	0.046	0.062	0.035
	400	0.030	0.036	0.043	0.171	0.170

To add perspective on the relative merits of method SACCP and the choice for the span, some additional simulation results are reported where the logistic regression model is correct and method SACCP is used to compute a confidence interval for $p(x)$ rather than $p_S(x)$. Reported in Table 3 are estimates of α for $\beta_1 = 0.1$, 0.3 and 0.5 and various choices for the span, h. For $\beta_1 = 0.1$, all of the estimates are less than 0.05, again the main limitation is that some estimates are less than 0.025 when the span is $h = 0.5$. Increasing the span to 0.75 improves the accuracy of the confidence intervals, but for $h = 1$, not shown, some estimates exceed 0.08. For $\beta_1 = 0.3$ and h=0.5, the results are fairly similar to those when $\beta_1 = 0.1$. But for $\beta_1 = 0.5$, using h=0.5 is unsatisfactory for certain values of x and $n = 200$, particularly for $n = 400$. Lowering the span to $h = 0.3$ yielded fairly accurate confidence intervals. So a rough characterization of SACCP is that generally $h = 0.5$ provides a reasonable choice for the span but exceptions occur when there is a relatively strong association and $n \geq 200$, in which case a smaller choice for h can be needed.

Illustrations

Data from two separate studies are used to illustrate method SACCP. The first illustration is based on data from the Well Elderly 2 study (Clark et al., 2011). The sample size is $n = 328$. The general goal was to assess the impact of an intervention program aimed at improving the health and wellbeing of older adults. Included were efforts aimed at understanding the nature of the association among various measures. Here the focus is on two measures taken after intervention: CESD, a measure of depressive symptoms, and LSIZ, a measure of life satisfaction. CESD scores greater than 15 are taken to be an indication of mild depression or worse. The goal here is to understand the association between LSIZ and the probability of a CESD score greater than 15. The upper left panel in Figure 3 shows the results based on the logistic regression model (method LT) and the upper right panel is the result using method SACCP.

For relatively high LSIZ scores, as LSIZ scores increase, the likelihood of having mild depression or worse decreases. However, for relatively low LSIZ scores, the two methods paint a decidedly different picture. For low LSIZ scores, the logistic regression model yields substantially shorter confidence intervals compared to SACCP. A histogram (not shown here) indicated LSIZ is skewed to the left with the bulk of the values greater than 10; values less than 10 are sparse. There are only 19 observations with an LSIZ score less than or equal to 8 and there are 32 observations less than or equal 10. This suggests inferences about the

regression line for LSIZ scores less than or equal 10 will be relatively imprecise. This is reflected by method SACCP in contrast to method LT. Also, SACCP suggests that the likelihood of CESD scores greater than 15 levels off substantially for LSIZ scores less than 10, but this is based on a relatively small amount of information.

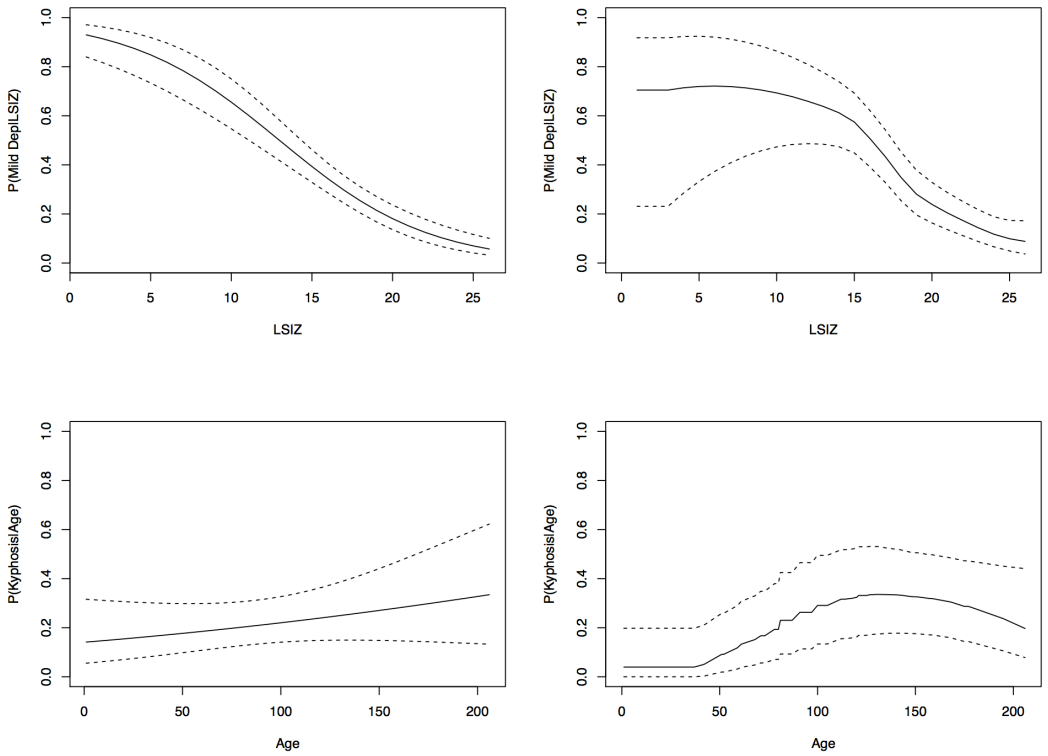

Figure 3. The top two panels show the results using method LT (SACCP) for the Well Elderly data. The bottom two panels show the results for the kyphosis data.

The second illustration considers the presence or absence of kyphosis, a postoperative spinal deformity. The sample size is $n = 81$. The focus is on the probability that kyphosis is present given the age of the patient in months (the software R contains the data in a built-in variable called kypho). The bottom left panel of Figure 3 shows the estimated regression line using method LT and the right panel is the estimate using SACCP. The two methods differ in fundamental ways. Method SACCP suggests for ages up to about 120 months, the probability of kyphosis increases more rapidly than indicated by method LT. Moreover, for ages

greater than 120, SACCP indicates the probability of kyphosis levels off and possibly decreases. The largest estimate based on SACCP is 0.36 and occurs for age 78 months. The largest estimate using LT is 0.33 and occurs at age 206 months.

Conclusion

Generally, method SACCP provides a more flexible approach to computing a confidence interval for $p(x)$ compared to method LT. Moreover, confidence intervals based on SACCP have the potential of being substantially more accurate. An advantage of the logistic regression model is that it can provide substantially shorter confidence intervals, but this can come at the expense of confidence intervals that might have very poor probability coverage, particularly when the sample size is large. But clearly there is room for improvement. For example, if there is a relatively strong association, there are situations where a span $h < 0.5$ would provide more accurate confidence intervals. And with a weak association, using $h > 0.5$ can yield reasonably accurate confidence intervals with shorter lengths compared to using $h = 0.5$. The best that can be done is to visually inspect the regression line using different values of h. If, for example, the association appears to be unusually strong, consider using $h = 0.4$ or even 0.3. This comes at a cost: wider confidence intervals. Yet another issue is making adjustments so that the simultaneous probability coverage is approximately equal to $1 - \alpha$. When dealing with a relatively small number of confidence intervals, there are well-known methods for dealing with this issue (e.g., Wilcox, 2017). An open issue is how to handle a large number of points.

An R function (`rplot.bin`) is available for applying method SACCP and is stored in the file `Rallfun-v35`, which can be downloaded from https://dornsife.usc.edu/labs/rwilcox/software/.

References

Agresti, A. & Coull, B. A. (1998). Approximate is better than `exact' for interval estimation of binomial proportions. *American Statistician, 52*(2), 119–126. https://doi.org/10.1080/00031305.1998.10480550

Blyth, C. R. (1986). Approximate binomial confidence limits. *Journal of the American Statistical Association, 81*(395), 843–855. https://doi.org/10.1080/01621459.1986.10478343

Brand, R. J., Pinnock, D. E. & Jackson, K. L. (1973). Large sample confidence bands for the logistic response curve and its inverse. *American Statistician, 27*(4), 157–160. https://doi.org/10.1080/00031305.1973.10479021

Brown, L. D., Cai, T. T. & DasGupta, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions. *Annals of Statistics, 30*(1), 160–201. https://doi.org/10.1214/aos/1015362189

Clark, F., Jackson, J., Carlson, M., Chou, C.-P., Cherry, B. J., Jordan-Marsh M., Knight, B. G., Mandel, D., Blanchard, J., Granger, D. A., Wilcox, R. R., Lai, M. Y., White, B., Hay, J., Lam, C., Marterella, A. & Azen, S. P. (2011). Effectiveness of a lifestyle intervention in promoting the well-being of independently living older people: results of the Well Elderly 2 Randomised Controlled Trial. *Journal of Epidemiology and Community Health, 66*(9), 782–790. https://doi.org/10.1136/jech.2009.099754

Clopper, C. & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika, 26*(4), 404–413. https://doi.org/10.1093/biomet/26.4.404

Copas, J. B. (1983). Plotting p against x. *Applied Statistics, 32*(1), 25–31. https://doi.org/10.2307/2348040

Fowlkes, E. B. (1987). Some diagnostics for binary logistic regression via smoothing. *Biometrika, 74*(3), 503–515. https://doi.org/10.1093/biomet/74.3.503

Hastie, T. J. & Tibshirani, R. J. (1990). *Generalized Additive Models*. New York: Chapman and Hall.

Hosmer, D. W., & Lemeshow, S. L. (1989). *Applied Logistic Regression*. New York: Wiley.

Hosmer, D. W., Hosmer, T., le Cessie, S. & Lemeshow, S. (1997). A comparison of goodness-of-fit tests for the logistic regression model. *Statistics in Medicine, 16*(9), 965–980. https://doi.org/10.1002/(sici)1097-0258(19970515)16:9<965::aid-sim509>3.0.co;2-o

Kay, R. & Little, S. (1987). Transformation of the explanatory variables in the logistic regression model for binary data. *Biometrika, 74*(3), 495–501. https://doi.org/10.1093/biomet/74.3.495

Khorasani, F. & Milliken, G.A. (1982). Simultaneous confidence bands for nonlinear regression models. *Communications in Statistics - Theory and Methods, 11*(11), 1241–1253. https://doi.org/10.1080/03610928208828308

Piegorsch, W. W. & Casella G. (1988). Confidence bands for logistic regression with restricted predictor variables. *Biometrics, 44*(3), 739–750. https://doi.org/10.2307/2531588

Pratt, J. W. (1968). A normal approximation for binomial, F, beta, and other common, related tail probabilities, I. *Journal of the American Statistical Association, 63*(324), 1457–1483. https://doi.org/10.1080/01621459.1968.10480939

Schilling, M. & Doi, J. (2014). A coverage probability approach to finding an optimal binomial confidence procedure. *American Statistician, 68*(3), 133–145. https://doi.org/10.1080/00031305.2014.899274

Tukey, J. W. & McLaughlin D. H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/Winsorization 1. *Sankhya A, 25*, 331–352.

Wilcox, R. R. (2017). *Introduction to Robust Estimation and Hypothesis Testing*. 4th Edition. San Diego, CA: Academic Press.