On iterative adjustment of responses for the reduction of bias in binary regression models

Ioannis Kosmidis

August 25, 2009

Abstract

The adjustment of the binomial data by small constants is a common practice in statistical modelling, for avoiding sparseness issues and, historically, for improving the asymptotic properties of the estimators. However, there are two main disadvantages with such practice: i) there is not a universal constant adjustment that results estimators with optimal asymptotic properties for all possible modelling settings, and ii) the resultant estimators are not invariant to the representation of the binomial data. In the current work, we present a parameter-dependent adjustment scheme which is applicable to binomial-response generalized linear models with arbitrary link functions. The adjustment scheme results by the expressions for the bias-reducing adjusted score functions in Kosmidis & Firth (2008, Biometrika) and thus its use guarantees estimators with second-order bias. Based on an appropriate expression of the adjusted data, a procedure for obtaining the bias-reduced estimates is developed which relies on the iterative adjustment of the binomial responses and totals using existing maximum likelihood implementations. Furthermore, it is shown that the bias-reduced estimator, like the maximum likelihood estimator, is invariant to the representation of the binomial data. A complete enumeration study is used to demonstrate the superior statistical properties of the bias-reduced estimator to the maximum likelihood estimator.
Keywords: bias reduction, adjusted responses, adjusted score functions

1 Introduction

In statistical modelling, the additive adjustment of binomial data by a constant a is a common practice for avoiding sparseness issues which may result to infinite maximum likelihood estimates and severe bias, or for improving the properties of the estimators, especially bias.

Consider independent binomial random variables $Y_1, \ldots, Y_n$ with totals $m_1, \ldots, m_n$ and probabilities $\pi_1, \ldots, \pi_n$, and the logistic regression model

$$\log\left(\frac{\pi_r}{1-\pi_r}\right) = \eta_r = \sum_{t=1}^{p} \beta_t x_{rt} \quad (r = 1, \ldots, n)\,, \tag{1}$$

with x_{rt} the (r,t)th component of an $n \times p$ design matrix X and with $\beta_1, \ldots, \beta_p$ unknown parameters (an intercept can be included in the model by setting the components of a column of X to one). Perhaps the most famous adjustment is the Haldane-Anscombe correction (Haldane, 1955; Anscombe, 1956), where $a = 1/2$ is appended to the binomial response and $2a$ to the binomial totals. Such an adjustment results in an estimator of the log-odds $\log\{\pi/(1-\pi)\}$ with bias of order $O(m^{-2})$ and has the further advantage that the resultant estimate is finite for every value of the response y. For the estimation of the parameters of a logistic regression model with $\eta_r = \beta_1 + \beta_2 x_r$ in (1), Hitchcock (1962) showed that the Haldane-Anscombe correction is not optimal in terms of the bias of the estimators and proposed $a = 1/4$ for $n = 3$ and no adjustment for $n > 3$. Hitchcock (1962) also noted that the first-order biases ($O(m^{-1})$ when $m = m_1 = \ldots = m_n$) depend on the parameter values (see, Gart and Zweifel, 1967, for a comparison of the above adjustments and some other adjustment schemes in terms of the bias of resultant log-odds estimators). In Gart et al. (1985), the Hitchcock (1962) proposal is verified for $n = 2, 3$ and 4 and refined for $n > 4$, demonstrating that for $n > 2$ there is not a universally optimal value of the constant a.

Clogg et al. (1991) presented a more sophisticated adjustment scheme for general logistic regressions where based on standard Bayesian arguments relating to the behaviour of the Jeffreys prior amongst every possible logistic regression, $a = p \sum_{r=1}^{n} y_r / (n \sum_{r=1}^{n} m_r)$ was appended to the binomial responses and p/n was to the binomial totals (see, also Rubin and Schenker, 1987). The stated aim in Clogg et al. (1991) was not bias reduction but rather an applicable method of eliminating the possibility of infinite maximum likelihood estimates for the many logistic regressions which were involved in the large application that was under consideration. The resultant estimator enjoys certain shrinkage properties which result in some reduction of the bias.

All the above adjustment schemes have the common property, and possibly were motivated by the fact, that a is constant with respect to the parameter vector $\beta = (\beta_1, \ldots, \beta_p)$ and thus estimation can be conveniently performed by the following procedure:

i) adjust the binomial data by a constant, and

ii) proceed with usual estimation methods, treating the adjusted responses as actual.

However, because the adjustments are constants, the resultant estimators are generally not invariant to different representations of the data (for example, aggregated and disaggregated view), a desirable invariance property that the maximum likelihood estimator has. Furthermore, the first-order bias of the maximum likelihood estimator for logistic regressions generally depends on the parameter values (see, Cordeiro and McCullagh, 1991, for explicit expressions) and thus, as is also amply evident from the studies in Hitchcock (1962) and Gart et al. (1985) there cannot be a universal constant a which always eliminates the first-order bias.

Firth (1993), in his study of bias-reducing adjusted score functions, presented a parameter-dependent adjustment scheme for logistic regressions which eliminates the first-order bias of the estimator. That adjustment scheme consists of the addition of half a leverage and a leverage to the binomial responses and totals, respectively.

In the current note, a parameter-dependent, bias-reducing adjustment scheme is proposed for the general class of binary-response generalized linear models with arbitrary link functions. The proposed adjustment scheme results naturally by the form of the bias-reducing adjusted score functions in Kosmidis and Firth (2008). An appropriate expression of the adjusted responses and totals is given so that the adjusted data mimic the range of the binomial responses and totals ($0 \le y_r \le m_r$). Based on that expression, an alternative to the modified iterative re-weighted least squares algorithm in Kosmidis and Firth (2008) is developed, where estimates with second-order bias can be obtained simply by the use of existing maximum likelihood implementations and appropriately adjusted responses and totals. Furthermore, it is shown that the adjustment schemes result in estimates that are invariant to the representation of the binomial data. A complete enumeration study is used to demonstrate the finiteness and shrinkage properties of the bias-reduced estimator as well as its better performance in terms of bias and mean squared error over the maximum likelihood estimator. Furthermore, the effect of bias reduction on the fitted probabilities is discussed.

2 Bias reducing adjustments to the score functions

Consider the same setup as for (1) but where the binomial probabilities are linked to the model parameters as

$$g(\pi_r) = \eta_r \quad (r = 1, \ldots, n) , \tag{2}$$

with $g(.)$ a monotone function from $[0, 1]$ to the real line. According to the results in Kosmidis & Firth (2008, Section 4.1), a second-order unbiased estimator of β can be obtained by the solutions of the adjusted score equations $U_t^* = 0$ $(t = 1, \ldots, p)$, with

$$U_t^* = \sum_{r=1}^{n} \frac{w_r}{d_r} \left(y_r + \frac{1}{2} h_r \frac{d_r'}{w_r} - m_r \pi_r \right) x_{rt} , \tag{3}$$

where $d_r = m_r \mathrm{d}\pi_r / \mathrm{d}\eta_r$, $d_r' = m_r \mathrm{d}^2\pi_r / \mathrm{d}\eta_r^2$, $w_r = d_r^2 / \{m_r \pi_r (1 - \pi_r)\}$ is the rth quadratic weight and h_r is the rth diagonal element of the hat matrix $H = X(X^T W X)^{-1} X^T W$, with W the diagonal matrix with non-zero elements w_r ($r = 1, \ldots, n$). The above adjusted score functions suggest appending $h_r d'_r/(2w_r)$ to the binomial response y_r ($r = 1, \ldots, n$). Kosmidis and Firth (2008) give the form of the adjusted responses for binomial-response generalized linear models for some well-known link functions.

3 Adjustment of the binomial responses and totals

3.1 An appropriate pseudo-data representation

A convenient way of solving the adjusted-score equations would be to use the adjusted responses in existing maximum likelihood implementations, iteratively. Nevertheless, a practical issue that can arise relates to the sign of $h_r d'_r/w_r$ (or simply the sign of d'_r) which can result in negative adjusted responses or adjusted responses greater than the binomial totals, violating the range of the actual data ($0 \leq y_r \leq m_r$). Fortunately, this issue can be resolved with simple algebraic manipulation.

Dropping the subject index r, a pseudo-data representation is defined as the pair $\{y^*, m^*\}$, with y^* the adjusted response and m^* the adjusted binomial total. By this definition, the apparent pseudo-data representation suggested by (3) is $\{y + hd'/(2w), m\}$. Nevertheless, the form of (3) suggests that there is a countable set of equivalent pseudo-data representations, equivalent in the sense that if the actual responses and totals are replaced with $\{y^*, m^*\}$ in the likelihood equations then the adjusted score equations result. Any pseudo-data representations in this set can be obtained from any other by the operations of adding and subtracting a quantity to either the adjusted responses or the adjusted totals, and of moving summands from the adjusted responses to the adjusted totals after division by $-\pi$.

Within this set of pseudo-data representations consider the ones which have the form

$$\left\{ y + h\frac{d'}{2w} + h\pi b \,, \quad m + hb \right\} , \tag{4}$$

with b some function of η. Substituting for w, because $0 \leq h \leq 1$ and $0 \leq y \leq m$, a sufficient condition for the adjusted responses and totals to mimic the range of the actual responses and totals (ie. $0 \leq y^* \leq m^*$ ($r = 1, \ldots, n$)) is that

$$b \geq md'(\pi - 1)/(2d^2)$$

and

$$b \geq md'\pi/(2d^2)$$

are satisfied simultaneously. Thus, because $0 \leq \pi \leq 1$, the requirement $0 \leq y^* \leq m^*$ can be met, for example, if $b = md'(\pi - 1)/(2d^2) + 1/2$ for $d' \leq 0$ and if $b = md'\pi/(2d^2) + 1/2$ for $d' > 0$. Hence, b can be set to $md'\left(\pi - I_{d' \leq 0}\right)/(2d^2) + 1/2$. Substituting in (4), we obtain the pseudo-data representation

$$\left\{ y + \frac{1}{2}h\pi \left(1 + \frac{md'}{d^2} I_{d'>0} \right) , \quad m + \frac{1}{2}h \left(1 + \frac{md'}{d^2} (\pi - I_{d' \leq 0}) \right) \right\} , \tag{5}$$

where I_E is 1 if E holds and 0 otherwise.

3.2 Local maximum likelihood fits on pseudo-data representations

In light of (5), the bias-reduced estimates can be obtained by an iterative adjustment procedure where the $(j + 1)th$ iteration is as follows:

i) Update to $\{y^*_{r,(j+1)}, m^*_{r,(j+1)}\}$ according to (5) ($r = 1, \ldots, n$) evaluating all the quantities involved at the estimates $\beta_{(j)}$ from the jth iteration.

ii) Use maximum likelihood to fit model (2) with responses $y^*_{r,(j+1)}$ and totals $m^*_{r,(j+1)}$ ($r = 1, \ldots, n$), using $\beta_{(j)}$ as starting value.

Table 1: asdasd

			Iteration 1		Iteration 2		Iteration 3		...	Iteration 11	
Log dose	y	m	$y^*_{(1)} - y$	$m^*_{(1)} - m$	$y^*_{(2)} - y$	$m^*_{(2)} - m$	$y^*_{(3)} - y$	$m^*_{(3)} - m$	...	$y^*_{(11)} - y$	$m^*_{(11)} - y$
1.691	6	59	0.13156	0.24572	0.13214	0.24644	0.13215	0.24645	...	0.13215	0.24645
1.724	13	60	0.14998	0.26399	0.14971	0.26309	0.14970	0.26306	...	0.14970	0.26306
1.755	18	62	0.14143	0.22935	0.14056	0.22772	0.14054	0.22768	...	0.14054	0.22768
1.784	28	56	0.09004	0.13721	0.08931	0.13611	0.08930	0.13609	...	0.08930	0.13609
1.811	52	63	0.10253	0.17514	0.10049	0.17154	0.10044	0.17145	...	0.10044	0.17145
1.837	53	59	0.16199	0.28238	0.15899	0.27705	0.15891	0.27691	...	0.15891	0.27691
1.861	61	62	0.14612	0.26159	0.14838	0.26526	0.14844	0.26536	...	0.14844	0.26536
1.884	60	60	0.03750	0.06974	0.04143	0.07690	0.04153	0.07709	...	0.04153	0.07709
Criterion	–		$O(10^{-6})$		$O(10^{-6})$		$O(10^{-7})$		...	$O(10^{-11})$	

If the maximum likelihood estimates are finite, they provide sufficiently good starting values for the above iteration. Otherwise, the iteration can start at the maximum likelihood estimates obtained after the addition of a constant $a > 0$ to the responses and $2a$ to the totals.

Furthermore, the condition $\sum_{t=1}^{p} |U_t^*(\beta_{(j+1)})| \leq \epsilon$, $\epsilon > 0$ can be used as a general convergence criterion of the procedure.

As an illustration of the above fitting procedure, consider a complementary log-log model for the beetle mortality data in Agresti (2002, Table 6.14). For the complementary log-log link $\pi = 1 - \exp(-\exp(\eta))$ and direct differentiation gives $d = m\exp\{\eta - \exp(\eta)\}$ and $d' = d\{1 - \exp(\eta)\}$. Substituting in (5) and expressing everything in terms of π, an appropriate pseudo-data representation for complementary log-log models has the form

$$\left\{ y + \frac{1}{2} h\pi \left(1 - \frac{1 + \log(1-\pi)}{(1-\pi)\log(1-\pi)} I_{\pi < c} \right) , \quad m + \frac{1}{2} h \left(1 - \frac{1 + \log(1-\pi)}{(1-\pi)\log(1-\pi)} (\pi - I_{\pi \geq c}) \right) \right\} ,$$

where $c = 1 - \exp(-1)$ is the probability for which $\eta = 0$. Table 1 gives the values of the additive adjustments to the responses and totals at each iteration of the fitting procedure in 5 significant places. Note that after the third iteration the changes to the adjustments appear after the sixth decimal. The bias-reduced estimates are -39.047 for the intercept and 21.748 for the logarithmic dose, with the corresponding maximum likelihood estimates being -39.522 and 22.015, respectively.

3.3 Invariance of the estimates to the structure of the data

It is always possible to represent Y_r as $\sum_{s=1}^{k_r} Z_{rs}$, where $Z_{r1}, \ldots, Z_{rk_r}$ are independent binomial random variables each with probability of success π_r and totals $l_{r1}, \ldots, l_{rk_r}$, respectively, with $m_r = \sum_{s=1}^{k_r} l_{rs}$ $(r = 1, \ldots, n)$. By this construction, in the presence of a covariate vector x_r for each observation y_r, the data for a binomial-response generalized linear model can be represented in equivalent ways, (y_r, m_r, x_r) $(r = 1, \ldots, n)$ and (z_{rs}, l_{rs}, x_r) $(r = 1, \ldots, n; s = 1, \ldots, k_r)$. The maximum likelihood estimator of the model parameters is invariant to the choice of either representation and, in contrast to constant adjustment schemes, the bias-reduced estimator also has the same invariance property.

To show this, denote z^*_{rs} and l^*_{rs} $(r = 1, \ldots, n; s = 1, \ldots, k_r)$, the adjusted responses and totals, respectively. Note that the bias-reducing pseudo-data representations have the generic form $\{z_{rs} + \tilde{h}_{rs} q_r, m_{rs} + \tilde{h}_{rs} v_r\}$ where $q_r \equiv q(\pi_r)$ and $v_r \equiv v(\pi_r)$ and $\tilde{h}_{rs}$ is the generic diagonal element of the hat matrix. Note that, q_r and v_r depend solely on π_r and, also, a simple calculation can show that $\sum_{s=1}^{k_r} \tilde{h}_{rs} = h_r$ $(r = 1, \ldots, n)$. Hence, $\sum_{s=1}^{k_r} z^*_{rs} = y^*_r$ and $\sum_{s=1}^{k_r} l^*_{rs} = m^*_r$ $(r = 1, \ldots, n)$ and because the adjusted score functions result by the replacement of the actual responses and totals by their adjusted versions, the bias-reduced estimates are invariant to the structure of the binomial data.

Table 2: Implied probabilities for $\beta_1 = -1$ and $\beta_2 = 1.5$ for the logistic, probit and complementary log-log link functions.

	π_1	π_2	π_3	π_4	π_5
logit	0.01799	0.07586	0.26894	0.62246	0.88080
probit	0.00003	0.00621	0.15866	0.69146	0.97725
cloglog	0.01815	0.07881	0.30780	0.80770	0.99938

4 Illustration of the properties of the bias-reduced estimator

To illustrate the properties of the bias-reduced estimator let $n = 5$ and $m_r = m$ $(r = 1, \dots, 5)$. Furthermore, consider the model (2) with $\eta_r = \beta_1 + \beta_2 x_r$ and let $x = (-2, -1, 0, 1, 2)$. For $m = 4, 8, 16$ and for the logit, probit and complementary log-log link functions, the bias and mean squared error of the bias-reduced estimator, as well as, the coverage of the nominally 95% Wald-type confidence interval were calculated through complete enumeration when the true parameter values are $\beta_1 = -1$ and $\beta_2 = 1.5$, which imply the extreme probability settings shown in Table 1. This calculation is possible, because for every data set and every link function, the bias-reduced estimates were finite. Nevertheless, the maximum likelihood estimator has at least one infinite component with positive probability. Thus the corresponding quantities for the maximum likelihood estimator are undefined and are only calculated conditionally on the finiteness of both its components.

The results are shown in Table 3. Direct comparison of the conditional moments and coverage for the maximum likelihood estimator with the corresponding unconditional quantities for the bias-reduced estimator is misleading, in the direction of favouring the method of maximum likelihood. Despite this reservation, according to the results in Table 3, in cases where the probability of infinite maximum likelihood estimates is small or moderate, the bias-reduced estimator has bias and mean squared error properties that are better, even, than the corresponding conditional quantities for maximum likelihood.

In Figure 1, using the results of the complete enumeration for the complementary log log link with $m = 4$, the fitted probabilities based on the bias-reduced estimates are plotted against the fitted probabilities based on the maximum likelihood estimates. The shrinkage of the former towards $1 - \exp(-1) \approx 0.632$ is apparent. Correspondingly, the bias-reduced estimates shrink towards the origin of the scale of the linear predictor relative to the maximum likelihood estimates. This behaviour is typical for binomial-response generalized linear models; for the probit and logit model the fitted probabilities shrink towards 0.5 (see, also, Kosmidis, 2007b).

From the results in Table 1, the coverage of the nominally 95% Wald-type confidence intervals for the bias-reduced estimator is poor in this setting. A reason for this is that the performance of the intervals is studied for extreme true parameter values; the finiteness and shrinkage properties of the bias-reduced estimator result in smaller estimated asymptotic standard errors (square roots of the diagonal of the Fisher information) so that the resultant Wald-type confidence intervals are short in length and do not cover extreme effects with sufficiently high probability. In contrast, for $m = 4$ and for true parameter values $\beta_1 = -1$ and $\beta_2 = 0.5$, the Wald-type confidence interval for the bias-reduced estimator of β_2 has better coverage behaviour with coverage 0.969 for the probit link and 0.971 for the complementary log log link. As the sample size increases the coverage tends to the nominal level.

5 Discussion

For binomial-response generalized linear models, it has been shown how second-order unbiased estimators can be resulted by iteratively adjusting the binomial responses and totals and using

Table 3: The results of the complete enumeration. The parenthesized probabilities refer to the event of encountering at least one infinite ML estimate for the corresponding m and link function. All the quantities for the ML estimator are calculated conditionally on the finiteness of its components.

Link	m	Parameter	Bias ($\times 10^2$)		MSE ($\times 10$)		Coverage	
			ML	BR	ML	BR	ML	BR
logit	4	β_1	-8.79	0.52	5.84	6.07	0.971	0.972
	(0.1621)	β_2	14.44	-0.13	3.62	4.73	0.960	0.939
	8	β_1	-12.94	-0.68	3.93	3.11	0.972	0.964
	(0.0171)	β_2	20.17	1.11	3.70	2.68	0.972	0.942
	16	β_1	-7.00	-0.19	1.75	1.42	0.961	0.957
	(0.0002)	β_2	10.55	0.29	1.59	1.16	0.960	0.948
probit	4	β_1	17.89	13.54	1.44	2.61	0.968	0.911
	(0.5475)	β_2	-18.84	-16.93	0.98	3.07	0.960	0.897
	8	β_1	0.80	3.24	1.07	1.82	0.964	0.938
	(0.2296)	β_2	6.08	-3.81	1.26	2.13	0.972	0.908
	16	β_1	-7.06	0.24	1.03	1.08	0.974	0.949
	(0.0411)	β_2	12.54	-0.17	1.39	1.22	0.973	0.933
cloglog	4	β_1	2.97	3.18	2.97	3.07	0.959	0.962
	(0.3732)	β_2	-2.93	-12.97	1.35	3.51	0.955	0.880
	8	β_1	-8.42	0.84	2.49	1.89	0.962	0.953
	(0.1000)	β_2	15.63	-5.40	2.33	2.36	0.972	0.906
	16	β_1	-6.45	0.17	1.32	0.98	0.964	0.957
	(0.0071)	β_2	13.13	-1.74	1.60	1.23	0.965	0.921

ML: maximum likelihood; BR: bias reduction.

existing maximum likelihood implementations.

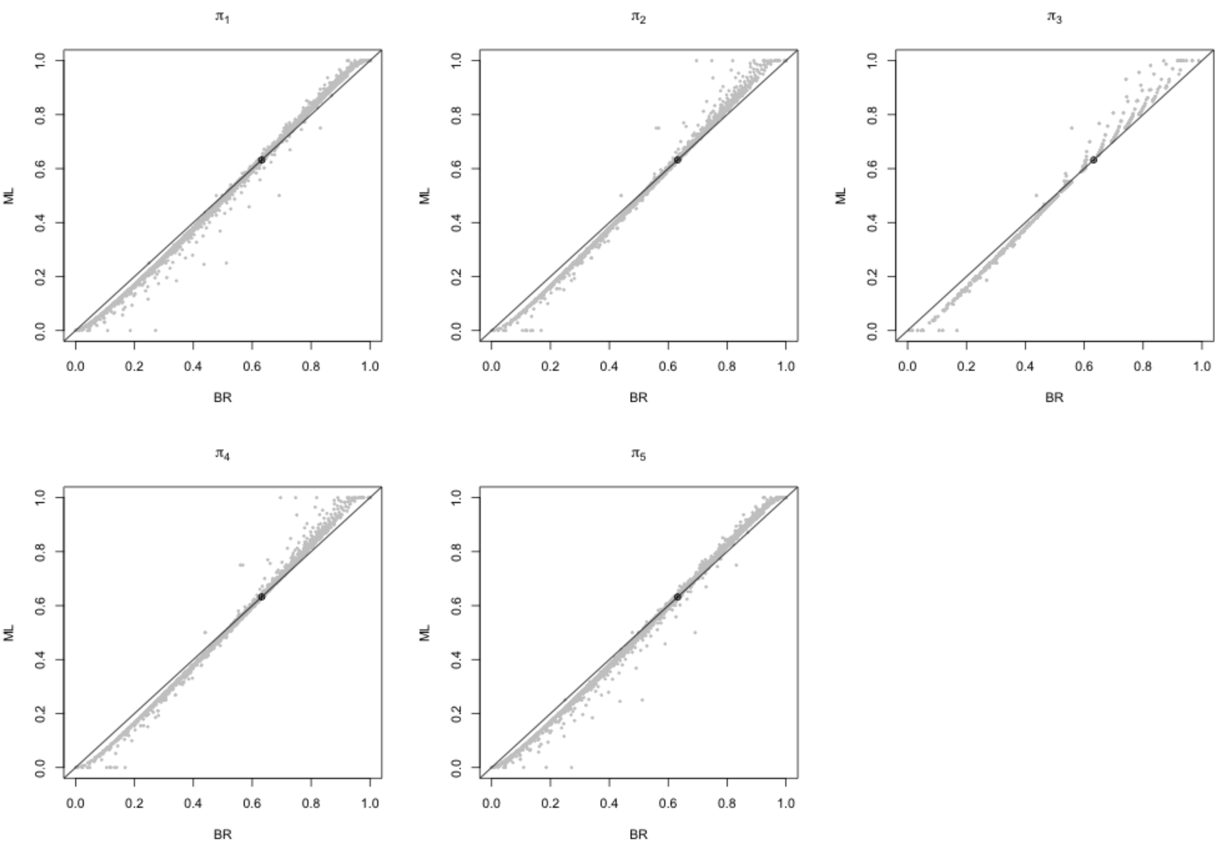
Like the maximum likelihood estimator, the bias-reduced estimator is invariant to the representation of the binomial data. In addition, contrastingly to the maximum likelihood estimates, as has been illustrated through complete enumeration, the bias-reduced estimates are always finite and because of their improved statistical properties, their routine use in applications is appealing.

Nevertheless, Wald-type approximate confidence intervals for the bias-reduced estimator can have bad coverage properties. In the case of logistic regression, the adjusted score functions correspond to the penalization of the likelihood by the Jeffreys invariant prior (Firth, 1993). Heinze and Schemper (2002) used this fact and illustrated that approximate confidence intervals based on the profiles of the penalized likelihood can have better coverage properties than Wald-type approximate confidence intervals. However, according to Kosmidis & Firth (2008, Theorem 1), the adjusted score functions for binomial-response generalized linear models with non-logistic link, do not generally admit a penalized likelihood interpretation. In those cases the adjusted-score statistic

$$U_t^*(\hat{\beta}_1, \ldots, \hat{\beta}_{t-1}, \beta_t, \hat{\beta}_{t+1}, \ldots, \hat{\beta}_p)^2 F^{tt}(\hat{\beta}_1, \ldots, \hat{\beta}_{t-1}, \beta_t, \hat{\beta}_{t+1}, \ldots, \hat{\beta}_p)$$

could be used for the construction of confidence intervals for the parameter β_t $(t = 1, \ldots, p)$. Here, $\hat{\beta}_u$ $(u = 1, \ldots, t-1, t+1, \ldots, p)$ are the bias-reduced estimates when the tth component of the parameter vector is fixed at β_t and F^{tt} is the (t,t)th component of the inverse Fisher information. Because $U_t^* = U_t + A_t$, where U_t is the tth component of the score vector and A_t is $O(1)$ as the sample size increases, the adjusted-score statistic is asymptotically distributed according to a chi-squared distribution with one degree of freedom.

Figure 1: Fitted probabilities based on the bias-reduced estimates against the fitted probabilities based on the maximum likelihood estimates for the cloglog link with $m = 4$. The marked point on the plots is (c, c), where $c = 1 - \exp(-1)$.



References

Agresti, A. (2002). *Categorical Data Analysis*. New York: Wiley.

Anscombe (1956). On estimating binomial response relations. *Biometrika 43*(3), 461–464.

Clogg, C. C., D. B. Rubin, N. Schenker, B. Schultz, and L. Weidman (1991). Multiple imputation of industry and occupation codes in census public-use samples using Bayesian logistic regression. *Journal of the American Statistical Association 86*, 68–78.

Cordeiro, G. M. and P. McCullagh (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society, Series B: Methodological 53*(3), 629–643.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika 80*(1), 27–38.

Gart, J. J., H. M. Pettigrew, and D. G. Thomas (1985). The effect of bias, variance estimation, skewness and kurtosis of the empirical logit on weighted least squares analyses. *Biometrika 72*, 179–190.

Gart, J. J. and J. R. Zweifel (1967, June). On the bias of various estimators of the logit and its variance with application to quantal bioassay. *Biometrika 54*(1), 181–187.

Haldane, J. (1955). The estimation of the logarithm of a ratio of frequencies. *Annals of Human Genetics 20*, 309–311.

Heinze, G. and M. Schemper (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine 21*, 2409–2419.

Hitchcock, S. E. (1962, June). A note on the estimation of parameters of the logistic function using the minimum logit χ^2 method. *Biometrika 49*(1), 250–252.

Kosmidis, I. and D. Firth (2008). Bias reduction in exponential family non-linear models. Technical Report 8-5, CRiSM working paper series, University of Warwick.

Rubin, D. B. and N. Schenker (1987). Logit-based interval estimation for binomial data using the Jeffreys prior. *Sociological Methodology 17*, 131–144.