Advances and Applications in Statistics
Volume 14, Number 2, 2010, Pages 101-116
Published Online: February 6, 2010
This paper is available online at http://www.pphmj.com

MAXIMUM LIKELIHOOD ESTIMATION WITH BINARY-DATA REGRESSION MODELS: SMALL-SAMPLE AND LARGE-SAMPLE FEATURES

ROLAND C. DEUTSCH*, JOHN M. GREGO[1], BRIAN HABING[1] and WALTER W. PIEGORSCH[2]

Department of Mathematics and Statistics
The University of North Carolina at Greensboro
Greensboro, NC 27402, U. S. A.
e-mail: rcdeutsc@uncg.edu

[1]Department of Statistics
University of South Carolina
Columbia, SC 29208, U. S. A.
e-mail: grego@stat.sc.edu; habing@stat.sc.edu

[2]Department of Mathematics and BIO5 Institute
University of Arizona, Tucson, AZ 85721, U. S. A.
e-mail: piegorsch@math.arizona.edu

Abstract

Many inferential procedures for generalized linear models (GLiMs) rely on the asymptotic normality of the maximum likelihood estimator (MLE). Fahrmeir and Kaufmann [5] present mild conditions under which the MLEs in GLiMs are asymptotically normal. Unfortunately, limited study has appeared for the special case of binomial response models beyond the familiar logit and probit links, with little results for more general links

2010 Mathematics Subject Classification: 62F25, 62F12, 62J12.

Keywords and phrases: binomial response models, maximum likelihood estimator, asymptotic normality, complementary log-log link, complementary log link.

*Corresponding Author

Received July 2, 2009

such as the complementary log-log link, and the less well-known complementary log link. We verify the asymptotic normality conditions of the MLEs for these models under the assumption of a fixed number of experimental groups and present a simple set of conditions for any twice-differentiable monotone link function. We also study the quality of the approximation for constructing asymptotic Wald confidence regions. Our results show that for small sample sizes with certain link functions the approximation can be problematic, especially for cases where the parameters are close to the boundary of the parameter space.

1. Introduction

Binomial response models are widely used when modeling binary response probabilities based on a set of (continuous and fixed) regressors. When employing such models, it is common to apply the maximum likelihood method for estimating parameters and base inference on the (supposed) asymptotic normality of the maximum likelihood estimator (MLE). Maximum likelihood estimation is discussed in Finney [7] for probit regression, in Hosmer and Lemeshow [11] for logistic regression models and in McCullagh and Nelder [13] and Agresti [1] for general binomial response models. These authors give details of how the estimation phase may be carried out, but only mention, and do not formally derive large sample properties such as asymptotic normality of the MLE. However, the latter three authors state that the asymptotic variance of the MLE is the inverse of the Fisher information matrix computed at the true parameter value, i.e., $\operatorname{Var}(\hat{\boldsymbol{\beta}}) = \mathcal{I}_N^{-1}(\boldsymbol{\beta}_0)$, and as such can be used to construct inferences on the vector of unknown parameters, $\boldsymbol{\beta}$. We detail these concepts in Section 2, below.

Work on asymptotic normality of the MLE in logistic regression can be found in McFadden [14] and in Nordberg [17]. McFadden presents regularity conditions for a multinomial response model when the logit link is used. Nordberg presents regularity conditions that assure asymptotic normality for the logit link in binomial response models and further verifies that his conditions are equivalent to those of McFadden [14]. However, the presented conditions are of a highly technical nature.

For dose-response models (one-regressor) Guess and Crump [8] consider maximum likelihood estimation using the less-common complementary log link (see Table 1, below) with linear predictor $\eta = \sum_{j=0}^{\infty} \beta_j d^j$ and q levels of the dose, d.

However, estimation is performed under the constraint $\beta_j \geq 0$ with strict inequality for at most $q + 1$ of the $\beta_j s$ and the others set to zero. In follow-up work, Crump et al. [3] discuss the asymptotic distribution of the MLE for constructing confidence intervals and conducting tests of hypotheses, while Guess and Crump [9] prove that the MLE is asymptotically normal in this setting as long as certain regularity conditions are satisfied.

For other link functions in binomial response models less work is evident. Haberman [10] presents highly technical conditions assuring existence and asymptotic normality of the MLE that require the number of parameters to grow as well. Based on the work of various former authors, Fahrmeir and Kaufmann [5] present mild, but still very technical, regularity conditions that assure existence and asymptotic normality of the MLE in the larger class of generalized linear models as defined by Nelder and Wedderburn [15]. They consider a very broad family of models that allow for many different forms of link function. Further, they also consider the special cases of binomial response models and models with compact regressors. In follow-up work, Fahrmeir and Kaufmann [6] further simplify their conditions for selected models, such as our binomial response with a logit or probit link, and prove existence and asymptotic normality for those models. Although less complex than the conditions presented by Haberman [10], their conditions are still quite technical (see Section 3 below). A very similar set of conditions implying asymptotic normality can be found in Silvapulle [21]. However, these conditions are used to establish convergence of certain test statistics rather than asymptotic normality of the MLE. Additionally, when considering other link functions besides the logit and probit link, e.g., the complementary log-log and complementary log links, we have found no work on clearly verifying these conditions or on presenting easily applicable conditions assuring asymptotic normality of the MLE. In this paper, we derive and verify practical regularity conditions for assuring asymptotic normality of the MLE for binomial response models under the assumption of a fixed number of experimental groups, based on the conditions from Fahrmeir and Kaufmann [5]. We place special emphasis on the complementary log-log and complementary log links. In Section 2, we introduce the class of binomial response models and link functions and derive likelihood-based quantities. In Section 3, we derive and verify conditions assuring asymptotic normality for the complementary log-log and complementary log links and also present a result for general monotone link functions. Finally, in Sections 4 and 5, we study the quality of the normal approximation when using finite samples for a variety of models involving one and two regressor variables. The conclusions from implications of these simulations are discussed in Section 6.

2. Binomial Response Models

Consider data $(\mathbf{Y}, \mathbf{X}_N)$ based on q experimental groups, where $\mathbf{Y}$ is a random response vector of size $N = \sum_{i=1}^{q} N_i$ and $\mathbf{X}_N$ is an $N \times p$ matrix $(N > p)$ of (fixed) regressors. The columns of $\mathbf{X}_N$ are assumed to be linearly independent (e.g., $\mathbf{X}_N$ has full rank), but may be functionally related. The individual responses follow a Bernoulli distribution, i.e.,

$$Y_{ij} \overset{ind.}{\sim} \mathbf{Ber}(\pi_i), \quad i = 1, \ldots, q; \quad j = 1, \ldots, N_i.$$

Further, let $Y_i = \sum_{j=1}^{N_i} Y_{ij}$ be the total number of responses at level i, and denote the response probability at that level as π_i. In order to connect the regressors to π_i, we use the linear predictor $\eta_i = \mathbf{x}_i'\boldsymbol{\beta}$, where $\boldsymbol{\beta} \in \boldsymbol{\mathcal{B}} \subseteq \mathbb{R}^p$ is the parameter vector and $\mathbf{x}_i \in \boldsymbol{\mathcal{X}} \subseteq \mathbb{R}^p$ is the corresponding vector of regressors. Two possible, relatively simple forms of the linear predictor are given by

$$\eta(x; \boldsymbol{\beta}) = \beta_0 + \beta_1 x, \text{ where } \mathbf{x}' = (1, x) \text{ and} \tag{1}$$

$$\eta(x, y; \boldsymbol{\beta}) = \beta_{00} + \beta_{10}x + \beta_{01}y + \beta_{11}xy, \text{ where } \mathbf{x}' = (1, x, y, xy). \tag{2}$$

The linear predictor is then associated with π using the link function $g(\pi) = \eta$, which is a strictly monotone function in π. The inverse link is then $g^{-1}(\eta) = \pi$. For the following, it will also be useful to consider this model in terms of the natural parameter, $\theta_i = \log\left(\frac{\pi_i}{1 - \pi_i}\right)$, rather than π_i. Obviously, the natural parameter space, Θ, of θ is the real line, $\mathbb{R}$. Further, when relating the linear predictor, η, to the natural parameter, we denote the inverse link by $h(\eta) = \theta$. Common link functions are given in Table 1.

Table 1. Considered links, inverse links and natural inverse links

	$g(\pi)$	$g^{-1}(\eta)$	$h(\eta)$
Logit	$\log\left(\frac{\pi}{1-\pi}\right)$	$(1+e^{-\eta})^{-1}$	η
Probit	$\Phi^{-1}(\pi)$	$\Phi(\eta)$	$\log\left(\frac{\Phi(\eta)}{1-\Phi(\eta)}\right)$
Comp. log-log	$\log(-\log(1-\pi))$	$1-\exp(e^{-\eta})$	$\log(\exp(e^{\eta})-1)$
Comp. log	$-\log(1-\pi)$	$1-e^{-\eta},\ \eta>0$	$\log(e^{\eta}-1),\ \eta>0$

As is well known [13], the mean and the variance of the responses Y_{ij} can now also be expressed in terms of the parameter vector $\boldsymbol{\beta}$ via $\mu_i(\boldsymbol{\beta}) = \pi_i = g^{-1}(\mathbf{x}_i'\boldsymbol{\beta})$ and $\sigma_i^2(\boldsymbol{\beta}) = \pi_i(1-\pi_i) = g^{-1}(\mathbf{x}_i'\boldsymbol{\beta})(1-g^{-1}(\mathbf{x}_i'\boldsymbol{\beta}))$, respectively.

The ML estimator $\hat{\boldsymbol{\beta}}$ of the parameter $\boldsymbol{\beta}$ can be obtained by maximizing the log-likelihood:

$$l_N(\boldsymbol{\beta};\mathbf{Y}) = \sum_{i=1}^{q}\sum_{j=1}^{N_i}(Y_{ij}\log(\pi_i) + (1-Y_{ij})\log(1-\pi_i))$$

$$= \sum_{i=1}^{q}[Y_i h(\mathbf{x}_i'\boldsymbol{\beta}) - N_i\log(1+\exp(h(\mathbf{x}_i'\boldsymbol{\beta})))].$$

Equivalently, we can find the MLE as the zero of the score function:

$$\mathbf{u}_N(\boldsymbol{\beta}) = \frac{\partial l_N(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{q}\mathbf{x}_i h_i'(\boldsymbol{\beta})(Y_i - N_i\mu_i(\boldsymbol{\beta})),$$

where $h_i'(\boldsymbol{\beta}) = \left.\frac{dh(\eta)}{d\eta}\right|_{\eta=\mathbf{x}_i'\boldsymbol{\beta}}$. Using similar notation the Fisher information is obtained as

$$\mathcal{I}_N(\boldsymbol{\beta}) = \mathrm{Cov}(\mathbf{u}_N(\boldsymbol{\beta})) = \sum_{i=1}^{q}\mathbf{x}_i\mathbf{x}_i'(h_i'(\boldsymbol{\beta}))^2 N_i\sigma_i^2(\boldsymbol{\beta}).$$

In the following sections, we shall establish that under certain regularity conditions the MLE $\hat{\boldsymbol{\beta}}$ is asymptotically normal. More precisely,

$$(\mathcal{I}_N^{1/2}(\boldsymbol{\beta}_0))'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}),$$

as $N \to \infty$ such that $\frac{N_i}{N} \to w_i > 0,\ i = 1, \ldots, q,$ where $\boldsymbol{\beta}_0$ denotes the true parameter vector.

3. Regularity Conditions for Asymptotic Normality

As noted above, Fahrmeir and Kaufmann [5] present regularity conditions for a very general class of generalized linear models. In this section, we adapt and verify these conditions for the binomial response model. For the following, let $\lambda_{\min}^{\mathbf{A}}$ denote the smallest eigenvalue of a matrix $\mathbf{A}$ and let $\mathbf{C}_N = \mathbf{X}_N'\mathbf{X}_N = \sum_{i=1}^{q} N_i \mathbf{x}_i \mathbf{x}_i'$. With this notation, the necessary regularity conditions to assure asymptotic normality of the MLE from Fahrmeir and Kaufmann [5] are:

(1) *Open parameter space*:

The parameter space $\mathcal{B}$ is open in $\mathbb{R}^p$ and convex.

(2) *Permissible linear predictor*:

$$\eta_i = \mathbf{x}_i'\boldsymbol{\beta} \in g((0, 1)); \quad i = 1, \ldots, q; \quad \forall \boldsymbol{\beta} \in \mathcal{B}.$$

(3) *Differentiability of link and inverse link*:

$g(\pi)$ and $h(\eta)$ are twice continuously differentiable and $|h'(\eta)| = \left|\frac{dh(\eta)}{d\eta}\right| > 0.$

(4) *Full rank of the cross product matrix*:

$\mathbf{C}_N$ has full rank for $N > N_0$.

(5) *Compact regressors*:

(5.1) The regressors $\mathbf{x}_i,\ i = 1, \ldots, q$ lie in a compact set $\mathcal{X}$ with $h(\mathbf{x}'\boldsymbol{\beta}) \in \Theta^0$ for all $\mathbf{x} \in \mathcal{X},\ \boldsymbol{\beta} \in \mathcal{B}$, where Θ^0 denotes the interior of the natural parameter space.

(5.2) $\lambda_{\min}^{\mathbf{C}_N}\mathbf{C}_N \to \infty.$

Deutsch [4, Section 4.3] shows that these conditions are specifically satisfied for binomial response models.

Note that instead of condition (5.2), Silvapulle [21] requires the distribution of the (random) regressors to be non-degenerate. Further, note that only conditions 1-3 rely upon the form of the link function. All other conditions are satisfied based on the fact that the regressors are finite. For binomial response models, the conditions of Fahrmeir and Kaufmann [5] are thus equivalent to:

Theorem 1 (Asymptotic normality of the MLE in binomial response models)**.** *Given* $\mathbf{x}_i \in \mathbb{R}^p$ *fixed,* $\boldsymbol{\beta}_0 \in \mathcal{B} \subseteq \mathbb{R}^p$, *the strictly monotone link function* $g(\pi)$ *and the model*

$$Y_{ij} \sim \mathbf{Ber}(g^{-1}(\mathbf{x}_i'\boldsymbol{\beta}_0)), \quad i = 1, \ldots, q, \quad j = 1, \ldots, N_i.$$

If

I. $g(\pi)$ *and* $g^{-1}(\eta)$ *are twice continuously differentiable,*

II. $\mathbf{X}_N = [\mathbf{1}_{N_i}\mathbf{x}_i']_{i=1,\ldots,q}$ *is of full rank p and*

III. $\mathcal{B}$ *open and convex,*

then

$$(\mathcal{I}_N^{1/2}(\boldsymbol{\beta}_0))'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}),$$

as $N \to \infty$ *such that* $\frac{N_i}{N} \to w_i > 0,\ i = 1, \ldots, q.$

The conditions in Theorem 1 may seem restrictive at first, but are met for a wide range of link functions. Conditions I and III are satisfied by the logit, probit, complementary log-log, complementary log, cauchit and tobit links, as well as for any other link based on a continuous distribution function. Also, the requirement for the full rank of the design matrix $\mathbf{X}_N$ is met in almost all regression settings, and can be readily verified.

4. Simulation Study for One-regressor Models

In addition to establishing the asymptotic normality of the MLE $\hat{\boldsymbol{\beta}}$, it is also of interest to assess the quality of this approximation when employing it in practice. Towards this end, we conducted a simulation study via the freeware package R [20].

The study was performed in two blocks, one involving models with only one regressor variable, the other involving models with two regressor variables.

The one-regressor models were taken from previous simulation studies in Nitcheva et al. [16] and Buckley and Piegorsch [2]. Six model parameterizations were considered, representing a range of commonly observed dose-response patterns; see Table 2. The parameters were obtained by using the four link functions from Table 1 with the linear predictor in (1). Notice that when using the complementary log link, models A and B represent situations where the parameter vector $\boldsymbol{\beta}$ is close to the boundary of the parameter space $\mathcal{B}$.

Table 2. One-regressor models: set up and parameters

Model		A	B	C	D	E	F
$P(Y=1 \mid x=0)$		0.01	0.01	0.10	0.05	0.30	0.10
$P(Y=1 \mid x=1)$		0.10	0.20	0.30	0.50	0.75	0.90
Logit	β_0	–4.5951	–4.5951	–2.1972	–2.9444	–0.8473	–2.1972
	β_1	2.3979	3.2088	1.3499	2.9444	1.9459	4.3944
Probit	β_0	–2.3263	–2.3263	–1.2816	–1.6449	–0.5244	–1.2816
	β_1	1.0448	1.4847	0.7572	1.6449	1.1989	2.5631
C. log-log	β_0	–4.6001	–4.6001	–2.2504	–2.9702	–1.0309	–2.2504
	β_1	2.3498	3.1002	1.2194	2.6037	1.3576	3.0844
Comp. log	β_0	0.0101	0.0101	0.1054	0.0513	0.3567	0.1054
	β_1	0.0953	0.2131	0.2513	0.6419	1.0296	2.1972

For our simulation study, the regressor values were taken as $x = 0, 0.25, 0.5, 1$, which corresponds to a common design in cancer risk experimentation [19]. For each model parameterization in Table 2, $n_{\text{sim}} = 2000$ binomial datasets were generated with $n = 25$ responses at each of the four regressor levels. From the generated data, the MLE $\hat{\boldsymbol{\beta}}$ was computed and saved. This procedure was repeated for $n = 50, 100, 300, 500, 1000, 5000, 10000$ at each of the regressor levels.

To study the quality of the normality approximation, we constructed (simultaneous) 95% Wald confidence regions for the full parameter vector, $\boldsymbol{\beta}$, and the (pointwise) slope parameter, β_1, at each repetition and studied their coverage probabilities. A simultaneous $(1-\alpha)$ Wald confidence region is the ellipsoid given by

$$\{\boldsymbol{\beta} : (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathcal{I}_N(\hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq \chi^2_\alpha(p)\}$$

[cf. 18, Section A.5.4]. Note that for $n_{\text{sim}} = 2000,$ the standard error of the estimated coverage near $1 - \alpha = 0.95$ is approximately $\sqrt{(0.95)(0.05)/2000} \approx 0.0049$ and does not exceed $\sqrt{(0.5)(0.5)/2000} \approx 0.0112.$

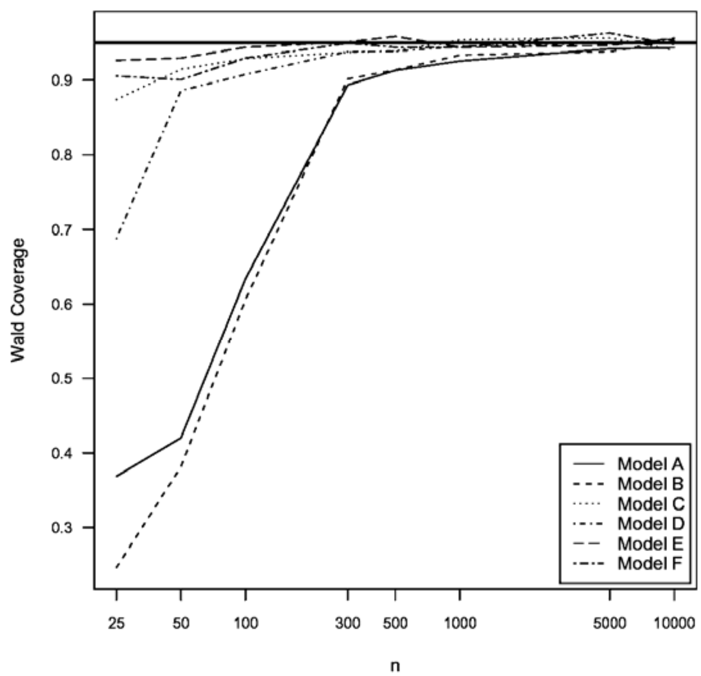

Figure 1. Simulated coverage for 95% Wald confidence regions on the full parameter vector when using the complementary-log link.

When inference is performed on the full parameter vector we find that the coverage of the Wald confidence region is close to the nominal level of 95% for each of the logit, probit or complementary log-log links. Our simulated coverage rates for all models and link functions ranged from 0.939 to 0.969 (results available from the authors), which is roughly within Monte Carlo sampling error of the nominal level. However, employing the complementary log link proves to be more problematic (see Figure 1). For models with very low background response (response at $x = 0$), namely models *A* and *B*, nominal coverage is not achieved until the individual sample sizes exceed $n = 5000$ and is extremely low for small sample sizes. Coverage performance improves as the background response increases: For model *D* with background response 0.05, the nominal coverage is achieved for sample sizes of 500 and above. Sample sizes of 100 and above are sufficient for models *C* and *F* with background response 0.1, while model *E* with a background response of 0.3 proves to be largely acceptable at any of the sample sizes studied.

When considering pointwise inferences only on the slope parameter, coverage performance improves over that of the joint confidence region. For sample sizes of 100 and above and for all links studied, empirical coverage rates ranged between 0.934 and 0.966 and were thus quite close to the nominal coverage level; however, at sample sizes of 50 and below, empirical coverage rates for models *A* and *B* when employing the logit, probit and complementary log-log links ranged between 0.96 and 0.982 and were thus somewhat conservative. The only concerns of any substance appeared again with the complementary log link: for models *A*, *B* and *C* with sample sizes of 50 and below, we generally observed coverage rates below 0.95, some as low as 0.923, suggesting slight under-coverage. (Results not shown; further details are available from the authors.)

5. Simulation Study for Two-regressor Models

We also considered models using two regressor variables as in (2). The simulation study was set up in similar fashion as for the one-regressor case. Five model parameterizations were studied, as summarized in Table 3. Models AB, CE, CF and EG were constructed from the one-regressor models in Table 1 (corresponding to each letter combination) and represent a variety of response surfaces. Model TNF was taken from a model fit of actual two-regressor data by Trinchieri et al. [22]. The predictor in (2) was then applied to obtain the $\boldsymbol{\beta}$ parameters. For both the regressors, x_1 and x_2 values were set to 0, 0.25, 0.5, 1. As with the one-regressor case, when using the complementary log link model AB represents a situation where $\boldsymbol{\beta}$ is close to the boundary of the parameter space $\mathcal{B}$.

Table 3. Two-regressor models: set up and parameters

Model		TNF	AB	CE	CF	EG
$P(Y=1\mid x_1=0,\ x_2=0)$		0.0550	0.0100	0.1000	0.1000	0.3000
$P(Y=1\mid x_1=1,\ x_2=0)$		0.5100	0.1000	0.3000	0.3000	0.7500
$P(Y=1\mid x_1=0,\ x_2=1)$		0.1950	0.2000	0.7500	0.9000	0.5000
$P(Y=1\mid x_1=1,\ x_2=1)$		0.9650	0.2727	0.9239	0.9720	0.9324
Logit	β_{00}	–2.8439	–4.5951	–2.1972	–2.1972	–0.8473
	β_{10}	2.8839	2.3979	1.3499	1.3499	1.9459
	β_{01}	1.4260	3.2088	3.2958	4.3944	0.8473
	β_{11}	1.8508	–1.9924	0.0477	0.0000	0.6782
Probit	β_{00}	–1.5982	–2.3263	–1.2816	–1.2816	–0.5244
	β_{10}	1.6233	1.0448	0.7572	0.7572	1.1989
	β_{01}	0.7386	1.4847	1.9560	2.5631	0.5244
	β_{11}	1.0483	–0.8078	0.0000	–0.1277	0.2950
C. log-log	β_{00}	–2.8723	–4.6001	–2.2504	–2.2504	–1.0309
	β_{10}	2.5345	2.3498	1.2194	1.2194	1.3576
	β_{01}	1.3440	3.1002	2.5770	3.0844	0.6644
	β_{11}	0.2035	–1.9941	–0.6001	–0.7794	0.0000
Comp. log	β_{00}	0.0566	0.0101	0.1054	0.1054	0.3567
	β_{10}	0.6568	0.0953	0.2513	0.2513	1.0296
	β_{01}	0.1603	0.2131	1.2809	2.1972	0.3365
	β_{11}	2.4787	0.0000	0.9378	1.0217	0.9713

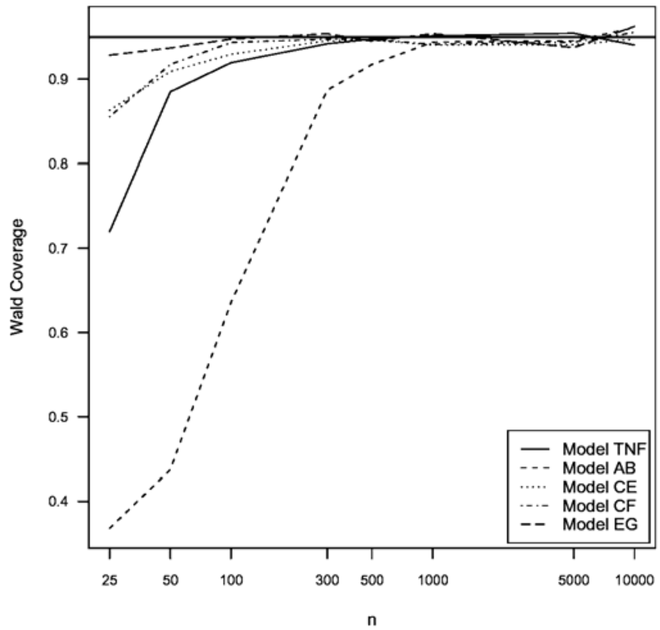

Figure 2. Simulated coverage for 95% Wald confidence regions on the full parameter vector when using the complementary-log link.

We studied the (simultaneous) coverage of 95% Wald confidence regions for the full parameter vector, $\boldsymbol{\beta}$, as well as for the non-intercept parameters, $(\beta_{10}, \beta_{01}, \beta_{11})$, and then pointwise for the cross-product parameter, β_{11}. Examining the simulated coverage rates for the full parameter vector, we found similar patterns as in the one-regressor case. When using the logit, probit and complementary log-log link, all simulated coverage rates over all studied sample sizes and models ranged from 0.94 to 0.9635 (result available from the authors). For the complementary log link, however, empirical coverage largely depended on the background response $(\text{at } x_1 = 0, x_2 = 0)$; see Figure 2. For model AB with a background response of 0.01 individual sample sizes of at least $n = 1000$ were required to achieve nominal coverage. For model TNF (background response 0.055) the required minimum individual sample size drops to $n = 500$. Models CE and CF both have a background response of 0.1 and require the individual sample sizes to be at $n = 300$ or above. Model EG (background response 0.3) displays hardly any coverage problems when using the complementary log link, with coverage initially hovering just below nominal coverage, which is then achieved for sample sizes of around $n = 100$.

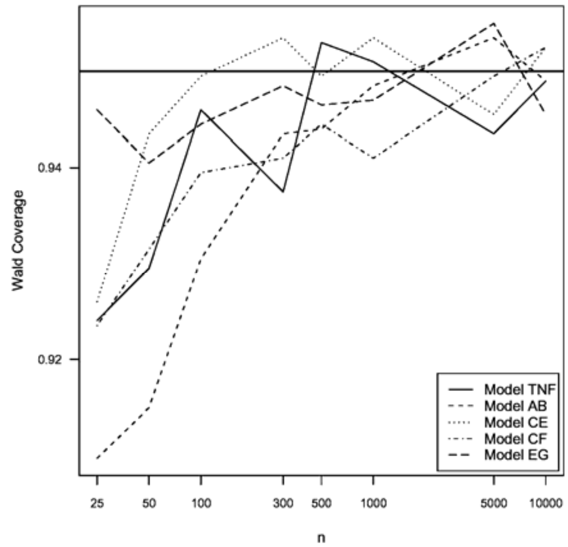

Figure 3. Simulated coverage for 95% simultaneous Wald confidence regions on the non-intercept parameters when using the complementary-log link.

When considering confidence regions on only the non-intercept parameters, coverage rates improve substantially. Our simulations continue to suggest that nominal coverage is achieved at all sample sizes studied when employing the logit, probit or complementary log-log links: empirical coverage rates ranged from 0.938 to 0.964 (results not shown). For the complementary log, however, empirical coverage drops below the nominal level at individual sample sizes less than $n = 300$ for model AB, and less than $n = 100$ for models TNF, CE and CF; see Figure 3. It is also interesting to observe that for all models the coverage rates are initially below nominal coverage and then converge to nominal levels as sample size increases. When pointwise inference on only the cross-product parameter β_{11} is considered, all links display acceptable coverage, with simulated coverage rates ranging from 0.934 to 0.964 (results available from the authors).

6. Discussion

We have presented easily verifiable conditions for the asymptotic normality of the MLE with binomial response models and have verified that those conditions satisfy the more general conditions in Fahrmeir and Kaufmann [5]. Through simulation, we further demonstrated that using this result to construct simultaneous confidence regions on $\boldsymbol{\beta}$ is generally acceptable when using the logit, probit or complementary log-log links. However, when using the complementary log link, our simulations show that the small-sample coverage critically depends on the background response. For models with a low background response (less than 0.1) nominal coverage might not be achieved even for individual sample sizes of 1000.

When considering inference on only the non-intercept parameters under a complementary log link, coverage rates improve and for most practical sample sizes, the normal approximation can be applied. Our results show that radical (anti-conservative) coverage is possible with the complementary log link, and that this may be due to failure of the intercept estimator to converge quickly to normality. This has some intuitive motivation: a low background response, which is essentially captured by the intercept, would translate into a potentially unstable boundary problem. This issue was recently discussed in a risk analytic scenario by Kopylev and Fox [12].

Concluding, we can say that for most practical situations the normality assumption for the MLE $\hat{\boldsymbol{\beta}}$ appears to be generally reasonable for the logit, probit and complementary log-log links, at least for the model parameterizations and sample sizes we studied. For the complementary log link, the approximation may be problematic with small sample sizes $(n < 100)$, however. For larger sample sizes the approximation should work reasonably well, assuming that $\boldsymbol{\beta}$ is well separated from the boundary of the parameter space $\mathcal{B}$. If some concern is evident over possible boundary-value concerns, then we should employ the approximation with the complementary log link only for much larger samples sizes, say, $n > 1000.$ A possible small-sample alternative in this case could be to employ bootstrap-based confidence statements; see Buckley and Piegorsch [2] and West et al. [23] for some preliminary results in this direction.

Acknowledgments

Thanks are due the editor and an anonymous referee for their support of this work, and to Brooke E. Buckley for her helpful comments during the development of the research. Dr. Deutsch's work was supported by new faculty Grant #123572 from the University of North Carolina at Greensboro. Dr. Piegorsch's work was supported by Grant #RD-832419 from the U.S. Environmental Protection Agency and Grant #R21-ES016791 from the U.S. National Institute of Environmental Health Sciences.

References

[1] A. Agresti, Categorical Data Analysis, 2nd ed., John Wiley & Sons, New York, 2002.

[2] B. E. Buckley and W. W. Piegorsch, Simultaneous confidence bands for Abbott-adjusted quantal response models in benchmark analysis, Statistical Methodology 5 (2008), 209-219.

[3] K. S. Crump, H. A. Guess and K. L. Deal, Confidence intervals and test of hypothesis concerning dose response relations inferred from animal carcinogenicity data, Biometrics 33 (1977), 437-451.

[4] R. C. Deutsch, Benchmark analysis for two predictor variables, Ph.D. Thesis, University of South Carolina, 2007.

[5] L. Fahrmeir and H. Kaufmann, Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models, Ann. Stat. 13 (1985), 342-368.

[6] L. Fahrmeir and H. Kaufmann, Asymptotic inference in discrete response models, Statist. Hefte (N.F.) 27 (1986), 179-205.

[7] D. Finney, Probit Analysis, 3rd ed., Cambridge University Press, Cambridge, 1971.

[8] H. A. Guess and K. S. Crump, Low-dose-rate extrapolation of data from animal carcinogenicity experiments - analysis of a new statistical technique, Math. Biosci. 32 (1976), 15-36.

[9] H. A. Guess and K. S. Crump, Maximum likelihood estimation of dose-response functions subject to absolutely monotonic constraints, Ann. Stat. 6 (1978), 101-111.

[10] S. J. Haberman, Maximum likelihood estimates in exponential response models, Ann. Stat. 5 (1977), 815-841.

[11] D. Hosmer and S. Lemeshow, Applied Logistic Regression, 2nd ed., John Wiley & Sons, New York, 2000.

[12] L. Kopylev and J. Fox, Parameters of a dose-response model are on the boundary: what happens with BMDL? Risk Anal. 29 (2009), 18-25.

[13] P. McCullagh and J. A. Nelder, Generalized Linear Models, 2nd ed., 37 of Monographs on Statistics and Applied Probability, Chapman & Hall, London, 1989.

[14] L. McFadden, Conditional Logit Analysis of Qualitative Choice Behavior, P. Zarembka ed., Frontiers in Econometrics, Academic, New York, 1974.

[15] J. A. Nelder and R. W. M. Wedderburn, Generalized linear models, J. Roy. Stat. Soc. A 135 (1972), 370-384.

[16] D. K. Nitcheva, W. W. Piegorsch, R. W. West and R. L. Kodell, Multiplicity-adjusted inferences in risk assessment: benchmark analysis with quantal response data, Biometrics 61 (2005), 277-286.

[17] L. Nordberg, Asymptotic normality of maximum likelihood estimators based on independent, unequally distributed observations in exponential family models, Scand. J. Stat. 7 (1980), 27-32.

[18] W. W. Piegorsch and A. J. Bailer, Analyzing Environmental Data, John Wiley & Sons, New York, 2005.

[19] C. J. Portier, Biostatistical issues in the design and analysis of animal carcinogenicity experiments, Environ. Health Persp. 102 (1994), 5-8.

[20] R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, 2005.

[21] M. J. Silvapulle, On tests against one-sided hypotheses in some generalized linear models, Biometrics 50 (1994), 853-858.

[22] G. Trinchieri, M. Kobayashi, M. Rosen, R. Loudon, M. Murphy and B. Perussia, Tumor necrosis factor and lymphotoxin induce differentiation of human myeloid cell lines in synergy with immune interferon, J. Exp. Med. 164 (1986), 1206-1225.

[23] R. W. West, D. K. Nitcheva and W. W. Piegorsch, Bootstrap methods for simultaneous benchmark analysis with quantal response data, Environ. and Ecol. Stat. 16 (to appear; doi: 10.1007/s10651-007-0073-5).