Statistica Sinica **15**(2005), 1081-1096

CONFIDENCE INTERVALS FOR THE MEAN VALUE OF RESPONSE FUNCTION IN GENERALIZED LINEAR MODELS

Jin Xu and Arjun K. Gupta

Bowling Green State University

Abstract: Confidence intervals for the mean value of the response function in generalized linear models are proposed to improve the accuracy of the approximation when the distribution of response is nonnormal and the sample size is moderate. The correction will give the approximation error up to order of $o(n^{-1/2})$ for the one-sided case and of $o(n^{-1})$ for the two-sided case. Monte Carlo studies are given to compare our results with the classical ones.

Key words and phrases: Confidence intervals, Edgeworth expansion, Skorohod representation, coverage probability, generalized linear models.

1. Introduction

Consider a generalized linear model (GLM) (McCullagh and Nelder (1989)) specified by univariate independent response variables and the canonical link function. We assume that y_i, $i = 1, \ldots, n$, are independent responses and each y_i has a density of the form

$$f(y_i;\theta_i) = \exp\{\, y_i\theta_i - b(\theta_i) + c(y_i)\}, \tag{1.1}$$

where $b(\cdot)$ and $c(\cdot)$ are known functions. The expected value of y is linked to a linear predictor $\eta = \boldsymbol{z}'\boldsymbol{\beta}$ by a known monotone function $g(\cdot)$ such that $\mathrm{E}[y|\boldsymbol{z}] = g(\eta)$. Here $\boldsymbol{z}$ and $\boldsymbol{\beta}$ are a p-dimensional known covariate and an unknown parameter, respectively. Under the canonical link, we have η as the natural parameter θ.

The problem of interest is to find a confidence interval for the mean response for a fixed covariate $\boldsymbol{z}$. Since $g(\cdot)$ is a known one-to-one function, this is equivalent to finding the confidence interval for the linear predictor η. One method constructs a two-sided confidence interval around the maximum likelihood estimate $\hat{\eta} = \boldsymbol{z}'\hat{\boldsymbol{\beta}}$ as

$$l_c = (\hat{\eta} - c\hat{\sigma}, \quad \hat{\eta} + c\hat{\sigma}), \tag{1.2}$$

where $\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimate of $\boldsymbol{\beta}$, $\hat{\sigma}^2 = \boldsymbol{z}'I^{-1}(\hat{\boldsymbol{\beta}})\boldsymbol{z}$ is the estimated asymptotic variance of $\hat{\eta}$ and $I(\hat{\boldsymbol{\beta}})$ is the Fisher information matrix evaluated at $\hat{\boldsymbol{\beta}}$. The two one-sided confidence intervals are constructed as

$$l_u = (-\infty, \quad \hat{\eta} + c_u\hat{\sigma}), \quad l_\ell = (\hat{\eta} - c_\ell\hat{\sigma}, \quad +\infty). \tag{1.3}$$

For a prespecified coverage probability $1-\alpha$, the constants c, c_u, and c_ℓ are determined by $1-\alpha = P(\eta \in l_c) = P(\eta \in l_u) = P(\eta \in l_\ell)$.

It is widely accepted that if the response variable is from a normal population or the sample size is sufficiently large, one can approximate the standardized Wald statistic $W_n = (\hat{\eta}-\eta)/\hat{\sigma}$ by a standard normal variable, hence the constants are obtained as $c = \Phi^{-1}(1-\alpha/2)$ and $c_u = c_\ell = \Phi^{-1}(1-\alpha)$, where $\Phi^{-1}(1-\alpha)$ is the upper α percentile of the standard normal distribution. The intervals given in (1.2) and (1.3), with these constants, are usually called Wald intervals in that they are obtained from the large sample Wald tests for $H_0 : \eta = (\geq, \leq)\, \eta_0$. It is worth noting that the Wald intervals actually have coverage error of order $n^{-1/2}$ in the one-sided case, and of order n^{-1} in the two-sided case. (Hall (1992, p.49))

However, there are many situations where the response variables are non-normal and only a moderate size sample is available due to practical constraints. Then coverage probability may no longer be as precise as the desired level. Moreover, when $g(\eta)$ is a constant function, Brown, Cai and DasGupta (2002, 2003) show that the Wald intervals have extremely poor performance in both discrete and continuous cases. Intensive numerical examples and theoretical explanations are given there to elucidate the striking erratic phenomenon. They recommend replacing the Wald intervals by score intervals, likelihood ratio intervals and Jeffrey's intervals (see the definition given in Brown, Cai and DasGupta (2003)) that show decisive improvement over the Wald intervals. But, in the general setting (1.1), when the mean value is a nonconstant function of the predictor, their proposed alternative intervals will not work as simply. The present paper attempts to fix this problem. We propose some improved confidence intervals based on the modified Wald statistic after making corrections for both nonnormality and the finiteness of the sample size. A parallel inference problem about testing under the same framework is discussed by Xu and Gupta (2003).

The main idea of the correction can be summarized as follows. First, we derive the Edgeworth expansion of the distribution of W_n that contains the skewness and the kurtosis effect from the underlying distribution. Second, a Skorohod representation is constructed to write W_n in terms of its limiting distribution. This is then followed by an inverse representation in terms of W_n to approximate its limiting distribution up to n^{-1}. We call this counterpart of Skorohod representation the modified Wald statistic. At last, we define the modified confidence intervals based on the new pivotal statistic.

Indeed, this idea can be traced back to Bartlett (1953). Johnson (1978) applied the same method to obtain modified t test and the confidence intervals for asymmetric populations. Abramovitch and Singh (1985) and Konishi (1991) also considered some general situations with connection to the bootstrap. More recently, Sun, Loader and McCormick (2000) used this Skorohod construction to improve the simultaneous confidence region for the mean response curve under generalized linear models. Our derivation of the Skorohod representation is similar to theirs in some parts, and differs in that we use the characteristic function.

The remainder of the paper is organized in the following way. Most details of the modified Wald intervals will be presented in Section 2. In Section 3, numerical examples are given to compare our improved confidence intervals with others. The Appendix contains the details of a proof.

2. Main Results

2.1. Edgeworth expansion of the pivotal statistic

To get at the distribution of the Wald statistic, we first obtain its Edgeworth expansion up to the order of n^{-1}.

We adopt the notation and follow the same route as did Sun, Loader and McCormick (2000). Given the density (1.1), the Fisher information matrix can be expressed as $I_n(\boldsymbol{\beta}) = \sum_{i=1}^n b''(\theta_i)\boldsymbol{z}_i\boldsymbol{z}_i'$. Let $A_n = I_n(\boldsymbol{\beta})/n$ be the rescaled Fisher information matrix and B_n be the upper Cholesky triangular matrix such that $B_n'B_n = A_n$. (Note that A_n is nonsingular almost surely.) Denote the log-likelihood function by $L(\boldsymbol{\beta}) = \sum_{i=1}^n \log f(y_i|\ \boldsymbol{\beta})$. We know that $\hat{\boldsymbol{\beta}}$ is the solution of the equation $\partial L(\boldsymbol{\beta})/\partial\boldsymbol{\beta} = 0$, i.e., $\sum_{i=1}^n (y_i - b'(\theta_i))\boldsymbol{z}_i = 0$. Assume that $\boldsymbol{\beta}_0$ is the true value of $\boldsymbol{\beta}$, let $\theta_{i0} = \boldsymbol{z}_i'\boldsymbol{\beta}_0$, and define $\boldsymbol{\psi}_n = \sum_{i=1}^n (y_i - b'(\theta_{i0}))\boldsymbol{z}_i/n$, $\boldsymbol{g}_n(\boldsymbol{\beta}) = \sum_{i=1}^n [b'(\theta_i) - b'(\theta_{i0})]\boldsymbol{z}_i/n$. Then $\boldsymbol{\beta}$ is the solution of $\boldsymbol{\psi}_n = \boldsymbol{g}_n(\boldsymbol{\beta})$.

Next, let $\boldsymbol{u}_i = B_n'^{-1}\boldsymbol{z}_i$ be the normalized covariate and define the normalized sum of independent vectors by $\boldsymbol{\xi} = \sum_{i=1}^n (y_i - b'(\theta_{i0}))\boldsymbol{u}_i/\sqrt{n}$. It is easily seen that $\boldsymbol{\xi} = \sqrt{n}B_n'^{-1}\boldsymbol{\psi}_n$ and $\mathrm{E}(\boldsymbol{\xi}) = \boldsymbol{0}$, $\mathrm{Cov}(\boldsymbol{\xi}) = I_p$. Then using the recursive method, we can expand the normalized MLE $\hat{\boldsymbol{\beta}}$ up to $\mathrm{O}_p(n^{-1})$ as

$$\sqrt{n}B_n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \boldsymbol{\xi} + \frac{1}{\sqrt{n}}\Big\{ -\frac{1}{2n}\sum_{i=1}^n b_i^{(3)}\boldsymbol{u}_i(\boldsymbol{u}_i'\boldsymbol{\xi})^2\Big\}$$
$$+\frac{1}{n}\Big\{ \frac{1}{2n^2}\sum_{i,j=1}^n b_i^{(3)}b_j^{(3)}\boldsymbol{u}_i(\boldsymbol{u}_i'\boldsymbol{\xi})(\boldsymbol{u}_i'\boldsymbol{u}_j)(\boldsymbol{u}_j'\boldsymbol{\xi})^2 - \frac{1}{6n}\sum_{i=1}^n b_i^{(4)}\boldsymbol{u}_i(\boldsymbol{u}_i'\boldsymbol{\xi})^3\Big\} + \mathrm{o}_p(n^{-1}).$$

Meanwhile, for a fixed covariate $\boldsymbol{z}$, let $\boldsymbol{\nu} = B_n'^{-1}\boldsymbol{z}/(\sqrt{n}\sigma_n)$ be the normalized covariate with $\|\ \boldsymbol{\nu}\| = 1$. Rewrite $W_n = [(\hat{\eta} - \eta)/\sigma_n]\cdot(\sigma_n/\hat{\sigma}_n)$, with $(\hat{\eta} - \eta)/\sigma_n = \langle\boldsymbol{\nu}, \sqrt{n}B_n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\rangle$ and $\sigma_n/\hat{\sigma}_n = [(\boldsymbol{z}'I_n^{-1}\boldsymbol{z})/(\boldsymbol{z}'\hat{I}_n^{-1}\boldsymbol{z})]^{1/2} = (\boldsymbol{\nu}B_n\hat{A}_n^{-1}B_n'\boldsymbol{\nu})^{-1/2}$. After some algebra, we finally obtain the expansion of W_n as

$$W_n = W_0 + \frac{1}{\sqrt{n}}W_1 + \frac{1}{n}W_2 + \mathrm{o}_p(n^{-1}), \tag{2.1}$$

where $W_0 = \boldsymbol{\nu}'\boldsymbol{\xi}$, and $W_1 = (1/2n)\sum_i b_i^{(3)}(\boldsymbol{u}_i'\boldsymbol{\nu})^2(\boldsymbol{u}_i'\boldsymbol{\xi})(\boldsymbol{\nu}'\boldsymbol{\xi}) - (1/2n)\sum_i b_i^{(3)}(\boldsymbol{u}_i'\boldsymbol{\nu})$ $(\boldsymbol{u}_i'\boldsymbol{\xi})^2$, and

$$\begin{aligned}
W_2 = &-\frac{1}{4n^2}\sum_{i,j} b_i^{(3)}b_j^{(3)}(\boldsymbol{u}_i'\boldsymbol{\nu})^2(\boldsymbol{u}_i'\boldsymbol{u}_j)(\boldsymbol{u}_j'\boldsymbol{\xi})^2(\boldsymbol{\nu}'\boldsymbol{\xi})\\
&-\frac{1}{2n^2}\sum_{i,j} b_i^{(3)}b_j^{(3)}(\boldsymbol{u}_i'\boldsymbol{\nu})(\boldsymbol{u}_j'\boldsymbol{\nu})(\boldsymbol{u}_i'\boldsymbol{\xi})(\boldsymbol{u}_i'\boldsymbol{u}_j)(\boldsymbol{u}_j'\boldsymbol{\xi})(\boldsymbol{\nu}'\boldsymbol{\xi})\\
&+\frac{3}{8n^2}\sum_{i,j} b_i^{(3)}b_j^{(3)}(\boldsymbol{u}_i'\boldsymbol{\nu})^2(\boldsymbol{u}_j'\boldsymbol{\nu})^2(\boldsymbol{u}_i'\boldsymbol{\xi})(\boldsymbol{u}_j'\boldsymbol{\xi})(\boldsymbol{\nu}'\boldsymbol{\xi})\\
&-\frac{1}{4n^2}\sum_{i,j} b_i^{(3)}b_j^{(3)}(\boldsymbol{u}_i'\boldsymbol{\nu})(\boldsymbol{u}_i'\boldsymbol{\xi})^2(\boldsymbol{u}_j'\boldsymbol{\nu})^2(\boldsymbol{u}_j'\boldsymbol{\xi})\\
&+\frac{1}{4n}\sum_{i} b_i^{(4)}(\boldsymbol{u}_i'\boldsymbol{\nu})^2(\boldsymbol{u}_i'\boldsymbol{\xi})^2(\boldsymbol{\nu}'\boldsymbol{\xi})\\
&-\frac{1}{6n}\sum_{i} b_i^{(4)}(\boldsymbol{u}_i'\boldsymbol{\nu})(\boldsymbol{u}_i'\boldsymbol{\xi})^3.
\end{aligned}$$

The last summations go through the index set $\{1, 2, \ldots, n\}$ for both i and j.

Remark 2.1. The above expansions of the normalized $\hat{\boldsymbol{\beta}}$ and W_n extend the results of Sun, Loader and McCormick (2000) by including the terms of order $O_p(n^{-1})$.

The computations for the first four cumulants of W_n, denoted by κ_j, $j = 1, 2, 3, 4$, follow from (2.1). It is found that κ_j is of order $n^{-(j-2)/2}$ and may be expanded as a power series in n^{-1} (see also Hall (1992, p.46)), i.e.,

$$\kappa_j = n^{-\frac{j-2}{2}}(\kappa_{j,1} + n^{-1}\kappa_{j,2} + n^{-2}\kappa_{j,3} + \cdots), \qquad j \geq 1, \tag{2.2}$$

with $\kappa_{1,1} = 0$, $\kappa_{2,1} = 1$ as desired. We omit the lengthy details and report the terms which serve as the coefficients in the Edgeworth expansion of the distribution up to n^{-1}. (Cf., Sun, Loader and McCormick (2000).) They are

$$\begin{aligned}
\kappa_{1,2} &= \frac{1}{2}C_1 - \frac{1}{2}C_2,\\
\kappa_{2,2} &= -3C_3 + \frac{1}{2}C_4 + \frac{1}{2}C_5 + \frac{7}{4}C_6 - \frac{1}{2}C_7 - \frac{1}{2}C_8 + C_9,\\
\kappa_{3,1} &= C_1,\\
\kappa_{4,1} &= -9C_3 + 6C_6 + 3C_9 + 3C_{10},
\end{aligned} \tag{2.3}$$

where

$$C_1 = \tfrac{1}{n}\sum b_i^{(3)}(\boldsymbol{u}_i'\boldsymbol{\nu})^3, \qquad C_3 = \tfrac{1}{n^2}\sum b_i^{(3)} b_j^{(3)}(\boldsymbol{u}_i'\boldsymbol{\nu})^2(\boldsymbol{u}_j'\boldsymbol{\nu})^2(\boldsymbol{u}_i'\boldsymbol{u}_j),$$
$$C_2 = \tfrac{1}{n}\sum b_i^{(3)}(\boldsymbol{u}_i'\nu)(\boldsymbol{u}_i'\boldsymbol{u}_i), \qquad C_4 = \tfrac{1}{n^2}\sum b_i^{(3)} b_j^{(3)}(\boldsymbol{u}_i'\boldsymbol{\nu})(\boldsymbol{u}_j'\boldsymbol{\nu})(\boldsymbol{u}_i'\boldsymbol{u}_j)^2,$$
$$C_8 = \tfrac{1}{n}\sum b_i^{(4)}(\boldsymbol{u}_i'\boldsymbol{\nu})^2(\boldsymbol{u}_i'\boldsymbol{u}_i), \qquad C_5 = \tfrac{1}{n^2}\sum b_i^{(3)} b_j^{(3)}(\boldsymbol{u}_i'\boldsymbol{\nu})^2(\boldsymbol{u}_i'\boldsymbol{u}_j)(\boldsymbol{u}_j'\boldsymbol{u}_j),$$
$$C_9 = \tfrac{1}{n}\sum b_i^{(4)}(\boldsymbol{u}_i'\boldsymbol{\nu})^4, \qquad C_6 = C_1^2,$$
$$C_{10} = \tfrac{1}{n}\sum b_i''^2(\boldsymbol{u}_i'\boldsymbol{\nu})^4, \qquad C_7 = C_1 \cdot C_2.$$

At last we get the Edgeworth expansion of the asymptotic distribution of W_n:

$$P(W_n \le x) = \Phi(x) + n^{-\frac{1}{2}} p_1(x)\phi(x) + n^{-1} p_2(x)\phi(x) + \mathrm{o}(n^{-1}), \tag{2.4}$$
$$p_1(x) = -[\kappa_{1,2} + \frac{1}{6}\kappa_{3,1}(x^2-1)],$$
$$p_2(x) = -x\Big[\frac{1}{2}(\kappa_{2,2}+\kappa_{1,2}^2)+\frac{1}{24}(\kappa_{4,1}+4\kappa_{1,2}\kappa_{3,1})(x^2-3) + \frac{1}{72}\kappa_{3,1}^2(x^4-10x^2+15)\Big], \tag{2.5}$$

where $\Phi(x)$ and $\phi(x)$ are the standard normal distribution function and density function, respectively. The details can be found in Hall (1992, Section 2.3).

Rremark 2.2. Strictly speaking, the Edgeworth expansion of the asymptotic distribution of W_n in (2.4) should include two additional oscillation terms, of order $n^{-1/2}$ and n^{-1}, when W_n posesses a lattice distribution. However this will rarely happen when the function $g(\eta)$ is nonconstant, even if the response variable is discrete. Hence we exclude them here. For more discussion on this issue, see Sun, Loader and McCormick (2000) and Bhattacharya and Rao (1976).

Rremark 2.3. In case a dispersion parameter, call it τ, is included in the original model, or the density of y_i is of the form

$$f(y_i;\theta_i,\tau) = \exp\{\ \frac{y_i\theta_i - b(\theta_i)}{\tau} + c(y_i,\tau)\}, \tag{2.6}$$

(2.4) holds with b_i'', $b_i^{(3)}$ and $b_i^{(4)}$ in the coefficients (2.3) replaced by b_i''/τ, $b_i^{(3)}/\tau$, $b_i^{(4)}/\tau$, respectively. The derivation proceeds in the same way with the above changes in the corresponding expressions. If τ is unknown, then we replace it by its consistent estimate.

2.2 Finiteness correction

The work of finding a transformation of W_n which absorbs the effect of high order cumulants of the population distribution and the finite sample size is achieved by the following proposition.

Proposition 2.1. *Suppose that the distribution of a random variable W can be expanded as*

$$P(W \le x) = \Phi(x) + n^{-\frac{1}{2}} p_1(x)\phi(x) + n^{-1} p_2(x)\phi(x) + \mathrm{o}(n^{-1}), \tag{2.7}$$

uniformly in x, where $p_1(x)$ and $p_2(x)$ are polynomials. Then we have

(i)
$$W \stackrel{d}{=} Z + n^{-\frac{1}{2}} q_1(Z) + n^{-1} q_2(Z), \tag{2.8}$$

$$\begin{aligned} q_1(x) &= -p_1(x), \\ q_2(x) &= -p_2(x) + p_1(x)p_1'(x) - \tfrac{x}{2} p_1^2(x); \end{aligned} \tag{2.9}$$

(ii)
$$W + n^{-\frac{1}{2}} h_1(W) + n^{-1} h_2(W) \stackrel{d}{=} Z, \tag{2.10}$$

$$\begin{aligned} h_1(x) &= -q_1(x), \\ h_2(x) &= -q_2(x) + q_1(x)q_1'(x). \end{aligned} \tag{2.11}$$

Additionally if we assume that $p_1(x)$ is even and $p_2(x)$ is odd, then we have

(iii)
$$|W| \stackrel{d}{=} |Z| - n^{-1} p_2(|Z|), \tag{2.12}$$

$$|W| + n^{-1} p_2(|W|) \stackrel{d}{=} |Z|. \tag{2.13}$$

Here $Z \sim N(0,1)$ and $U \stackrel{d}{=} V$ means $|P(U \le x) - P(V \le x)| = \mathrm{o}(n^{-1})$ uniformly in x.

The proof is deferred to the Appendix.

Rremark 2.4. The additional assumption about the types of symmetry of functions p_1 and p_2 is crucial to get the simple form of the Skorohod representation in (iii) of Proposition 2.1. Fortunately in most circumstances, these symmetries apply. (See Hall (1992), Appendix IV, for a counterexample.)

Rremark 2.5. It is easy to verify that the Skorohod representations in (2.8) and (2.12) can be extended to any order $n^{-k/2}$ as long as the corresponding expansion exists. A multi-dimensional generalization has also been studied. (See Xu and Gupta (2004).)

Now we are ready to construct two new pivotal statistics. Define

$$W_n^{(1)} = W_n + n^{-\frac{1}{2}} p_1(W_n) + n^{-1}\Big(p_2(W_n) + \frac{W_n}{2} p_1^2(W_n)\Big), \tag{2.14}$$

$$W_n^{(2)} = |W_n| + n^{-1} p_2(|W_n|), \tag{2.15}$$

where p_1 and p_2 are given in (2.5). By Proposition 2.1 we have $W_n^{(1)} \stackrel{d}{=} Z$ and $W_n^{(2)} \stackrel{d}{=} |Z|$ with the approximation error of order $\mathrm{o}(n^{-1})$. Then two modified one-sided confidence intervals and the modified two-sided confidence interval based on $W_n^{(1)}$ and $W_n^{(2)}$ for η are constructed as $l_u^* = \{\eta : W_n^{(1)} \geq -c_u\}$, $l_\ell^* = \{\eta : W_n^{(1)} \leq c_\ell\}$ and $l_c^* = \{\eta : W_n^{(2)} \leq c\}$. Obviously one can also construct a two-sided confidence interval from $W_n^{(1)}$ as $\tilde{l}_c = \{\eta : |W_n^{(1)}| \leq c\}$, but this will result in a coverage error of order n^{-1} in practice.

Ideally, the coverage probability of modified confidence intervals would be of order $\mathrm{o}(n^{-1})$ away from the nominal confidence coefficient $1-\alpha$. However, in practice, p_1 and p_2 have to be replaced by their estimated versions $\hat{p}_1$ and $\hat{p}_2$, obtained by substituting coefficients by their consistent estimates. In other words, we use $\hat{l}_u^* = \{\eta : \widehat{W}_n^{(1)} \geq -c_u\}$, $\hat{l}_\ell^* = \{\eta : \widehat{W}_n^{(1)} \leq c_\ell\}$ and $\hat{l}_c^* = \{\eta : \widehat{W}_n^{(2)} \leq c\}$ in applications. This might cause changes in the corresponding coverage probabilities. For instance, usually we have $\hat{p}_j = p_j + \mathrm{O}_p(n^{-1/2})$. By approximating the roots of $W_n^{(1)} = -c_u$ and $W_n^{(2)} = c$, we may find that

$$l_u^* = \{\eta : W_n \geq -[c_u + n^{-\frac{1}{2}} s_1(c_u) + n^{-1} s_2(c_u)] + \mathrm{o}(n^{-1})\},$$

where $s_1(x) = p_1(x)$, $s_2(x) = -p_2(x) + p_1(x)p_1'(x) - (x/2)p_1^2(x)$, and

$$l_c^* = \{\eta : |W_n| \leq c - n^{-1} p_2(c) + \mathrm{o}(n^{-1})\}. \tag{2.16}$$

By Proposition 3.1 of Hall (1992, p.102), we find that

$$P(\eta \in \hat{l}_u^*) = 1 - \alpha + n^{-1} b_\alpha \phi(c_u) + \mathrm{o}(n^{-1}), \tag{2.17}$$

where $b_\alpha = [(1/2)(C_9 - C_8) + (1/6)(c_u^2 - 1)C_9]c_u$. The first order correction yields an extra term $-n^{-1}s_2(c_u)$ on the RHS of (2.17). Further inspection shows that the magnitude of b_α is generally small. (Furthmore, note that $C_9 \leq C_8 \leq (1/n)\sum b_i^{(4)}(\boldsymbol{u}_i'\boldsymbol{u}_i)^2$ and the last quantity of the inequality is a certain version of kurtosis of the normalized vector $\boldsymbol{\xi}$.) It can also be seen from the simulation in the next section that the second order correction is superior to the first order correction in most cases. On the other hand, we have

$$P(\eta \in \hat{l}_c^*) = 1 - \alpha + \mathrm{o}(n^{-1}), \tag{2.18}$$

ensuring that the modified two-sided confidence interval behaves well.

Consider assessing the confidence intervals by their expected length. Let us just focus on the two-sided case. From (2.16) we can see that the expected length of the modified two-sided confidence interval l_c^* is $2[c - n^{-1}p_2(c)]\,\mathrm{E}(\hat{\sigma})$, compared to $2c\,\mathrm{E}(\hat{\sigma})$ for the Wald interval l_c in (1.2). Unfortunately we cannot be sure that $p_2(c)$ stays positive or, equivalently, that the modified confidence interval has a shorter expected length, although it seems that in most cases it does. Neither can we guarantee that the modified Wald interval has the shortest expected length among all confidence intervals of the same size, especially when

W_n has a skewed distribution. Further investigation is needed. Nevertheless the proposed confidence intervals satisfy the precision requirement, the primary objective of this paper.

3. Simulations and Application

In this section, we present the Monte Carlo investigation of three examples to demonstrate the performance of our proposed confidence intervals in comparison with existing methods. This is followed by an application to data.

Example 3.1. [Gamma response] Assume that $y_i, i = 1, \ldots, n$, are distributed according to Gamma(ν, μ_i) with densities

$$\frac{1}{\Gamma(\nu)}(\frac{\nu}{\mu_i})^{\nu} y_i^{\nu-1} \exp\{-\frac{\nu}{\mu_i} y_i\},$$

where the dispersion parameter $a(\tau) = \nu^{-1}$ is assumed to be 0.5 (so that the skewness of y_i is $\sqrt{2}$). The canonical link function is the reciprocal function associated with a linear predictor given by $\mu = \eta^{-1} = (5 + 2z)^{-1}$. We compare the coverage probabilities of the two one-sided confidence intervals (l_u vs l_u^*) and two two-sided confidence intervals (l_c vs l_c^*) at prescribed confidence coefficients.

The Monte Carlo study proceeds as follows.

1. Choose a sequence of values of the covariates randomly. (Make sure that each μ_i is positive, here we let z_i be Uniform(0,1).) Generate one observation from Gamma(ν, μ_i) for each μ_i.
2. Use Newton-Raphson or the iterative weighted least square method to obtain the MLE $\hat{\boldsymbol{\beta}}$.
3. Choose a number randomly from $(0, 1)$ as the covariate z. Compute W_n, $W_n^{(1)}$, $W_n^{(2)}$ and the estimated version $\widehat{W}_n^{(1)}$, $\widehat{W}_n^{(2)}$ from the assumed model and the sample, respectively.
4. Repeat 1∼3 sufficiently many times to obtain the empirical coefficient of the confidence intervals. (Here we set the simulation size to be 10,000.)

The results are displayed in Figure 3.1, where the symbols 'o', '+', '$*$' indicate, respectively, the coverage probabilities of confidence intervals constructed from W_n, $W_n^{(1)}$, $\widehat{W}_n^{(1)}$ in the one-sided case, and from $|W_n|$, $W_n^{(2)}$, $\widehat{W}_n^{(2)}$ in the two-sided case. Some discussion follows Example 3.3.

Example 3.2. [Poisson response] Assume the response variable of the GLM is Poisson with a *log-quadratic* link function, i.e., $\log \mathrm{E}[y] = -1 + 4x - 6x^2$, $x \in (0, 1)$. The simulation is carried out in a similar fashion as Example 3.1. We show the comparison of coverage probabilities for two types of confidence intervals in Figure 3.2. We also examine the global distributions of the pivotal statistic W_n and the new pivotal statistic $W_n^{(1)}$ after the finiteness correction by the sample relative frequency distributions, see Figure 3.3.

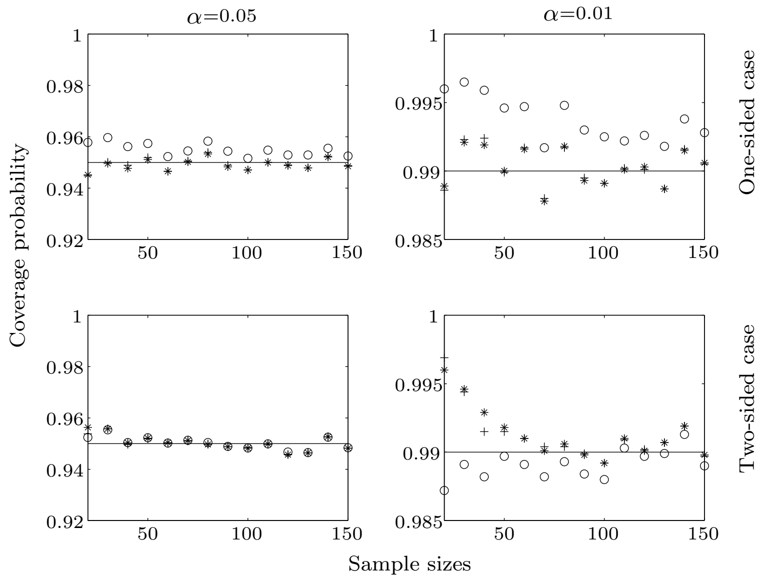

Figure 3.1. Coverage probabilities of confidence intervals of a Gamma reciprocal-linear regression. 'o'$\sim W_n(|W_n|)$, '+'$\sim W_n^{(1)}(W_n^{(2)})$, '$*$'$\sim \widehat{W}_n^{(1)}(\widehat{W}_n^{(2)})$ in one-sided (two-sided) case, respectively.

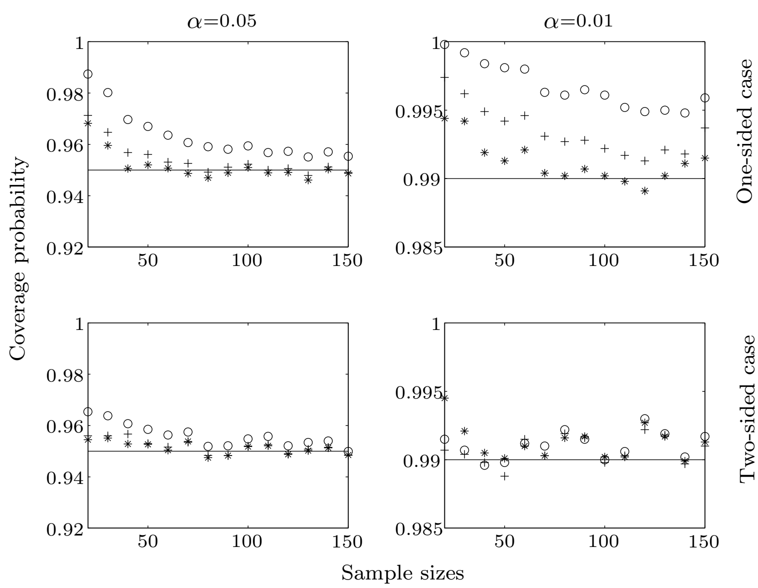

Figure 3.2. Coverage probabilities of confidence intervals of a Poisson log-quadratic regression. 'o'$\sim W_n$ $(|W_n|)$, '+'$\sim W_n^{(1)}$ $(W_n^{(2)})$, '$*$'$\sim \widehat{W}_n^{(1)}$ $(\widehat{W}_n^{(2)})$ in one-sided (two-sided) case, respectively.

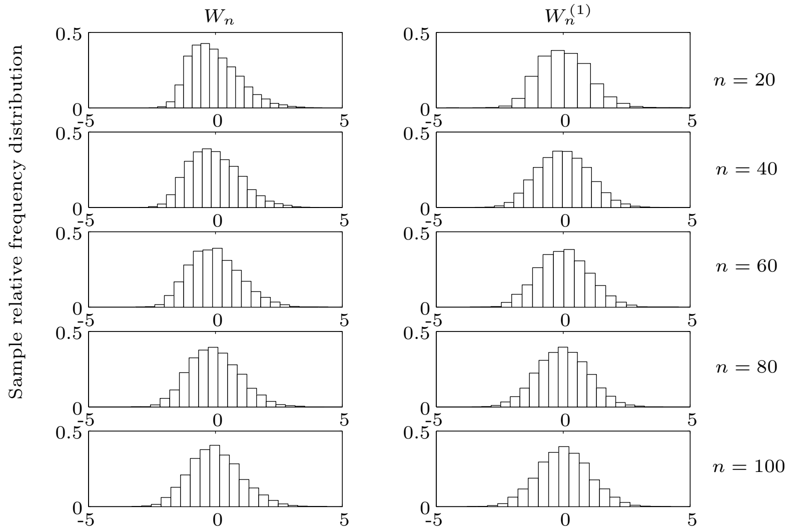

Figure 3.3. Comparison of the sample distributions of W_n and $W_n^{(1)}$ of a Poisson log-quadratic regression model.

Another comparison between the first order correction (up to $n^{-1/2}$) and the second order correction (up to n^{-1}) for the one-sided case is shown in Figure 3.4, where we plot the coverage probabilities of confidence intervals constructed by $\{\eta : W_n + n^{-1/2}\hat{p}_1(W_n) \geq -c_u\}$ and by $\hat{l}_u^* = \{\eta : \widehat{W}_n^{(1)} \geq -c_u\}$.

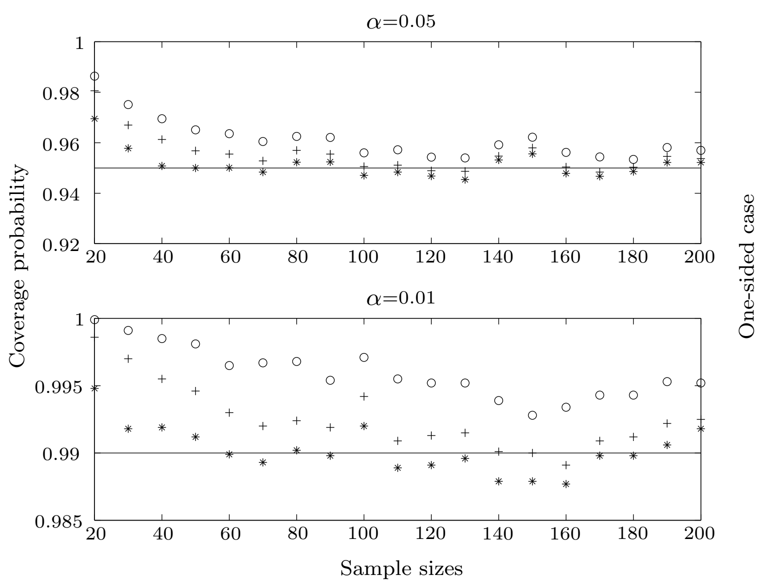

Figure 3.4. Coverage probabilities of confidence intervals of a Poisson log-quadratic regression. 'o'$\sim W_n$, '+'$\sim W_n + n^{-1/2}\hat{p}_1(W_n)$, '*'$\sim \widehat{W}_n^{(1)}$ in one-sided case, respectively.

Example 3.3 [Bernoulli response] Consider a Bernoulli random variable with a *logit-linear* link function specified by $\operatorname{logit} \mathrm{E}[y] = -1.5 + 3x$, $x \in (0, 1)$. On observing independent 0–1 responses y_i, $i = 1, 2, \ldots, n$, and associated covariates x_i, we compare the coverage probabilities of one-sided and two-sided confidence intervals for a randomly selected linear predictor. The results are shown in Figure 3.5, with the sample size ranging from 20 to 150 (with step 10) and 150 to 500 (with step 25).

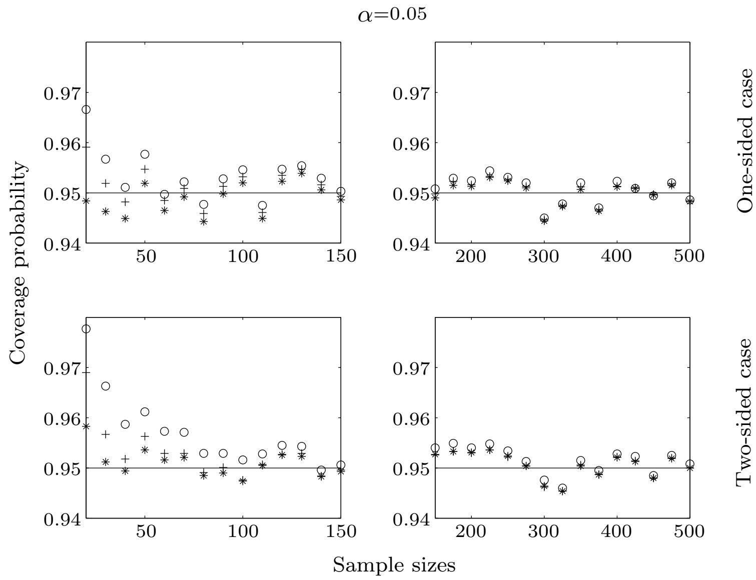

Figure 3.5. Coverage probabilities of confidence intervals of a Bernoulli logit-linear regression. 'o'$\sim W_n$ ($|W_n|$), '+'$\sim W_n^{(1)}$ ($W_n^{(2)}$), '$*$'$\sim \widehat{W}_n^{(1)}$ ($\widehat{W}_n^{(2)}$) in one-sided (two-sided) case, respectively.

To summarize the overall performance of the proposed confidence intervals from the Monte Carlo study, the improvement over the classical method is significant. To be more specific, in the one-sided cases of Gamma and Poisson responses, the correction to the skewness (Figure 3.3) and second order correction to the kurtosis and second effect of skewness (Figure 3.4) are self-evident. In the two-sided case, the correction for the Bernoulli response is quite remarkable when $n \leq 150$ (Figure 3.5), although we expect a minor change in the precision since the correction made is of the order of n^{-1}. On the other hand, large sample theory shows its domination in both cases as the sample size gets large enough. Another point we want to make is that the confidence intervals from the new pivotal statistics with estimated coefficients are almost indistinguishable from those of the true model, as we can see from Figure 3.1, Figure 3.2 and Figure 3.5

that $W_n^{(1)}$ and $\widehat{W}_n^{(1)}$, or $W_n^{(2)}$ and $\widehat{W}_n^{(2)}$, or '+' and '$*$' are close. This endorses the observations about the coverage probabilities (2.17) and (2.18) at the end of Section 2.

Finally we recommend to practitioners the following situations in which to apply the correction to the three different models (Table 3.1). With large sample size, say 300 or over, one may just use the first order correction to avoid extra calculation while still maintaining the claimed significance level.

Table 3.1. Recommended situations.

Model	Sample size	Type
Gamma	≤ 50	one-sided
Poisson	≤ 100	one & two-sided
Bernoulli	≤ 150	two-sided

We close this section by applying our method to data.

Table 3.2. Potato flour dilution.

Dilution	Spore growth		Proportion
(g/100ml)	No. of Plates	No. of Positve	of positive plates
1/128	5	0	0.0
1/64	5	0	0.0
1/32	5	2	0.4
1/16	5	2	0.4
1/8	5	3	0.6
1/4	5	4	0.8
1/2	5	5	1.0
1	5	5	1.0
2	5	5	1.0
4	5	5	1.0

Example 3.4 [Potato flour dilutions] A data set from Fisher and Yates (1963) records the growth of spores in ten dilutions of a suspension of a potato flour (Table 3.2). For each level of dilution, five plates are tested for positive growth. There are reasons to model this number (the response variable) by a binomial distribution associated with a *logit-linear* regression (McCulloch and Searle (2001, p.144)). To be more specific, let y_i, $i = 1, 2, \ldots, 10$, be the number of plates with positive growth, and x_i be the logarithm of the dilution in the corresponding level; we then assume that

$$\begin{cases} y_i \sim \text{indep. Binomial } (5, p_i), \\ \mathrm{E}[\, y_i] = 5\, p_i = 5\, \frac{1}{1+e^{-(\beta_0+\beta_1 x_i)}}\,. \end{cases}$$

We apply the proposed method to construct pointwise 95% confidence intervals for the positive growth proportion. It is seen in Figure 3.6 that the proposed confidence intervals (dot-dash lines) are narrower than those from the naive method (dotted lines). In other words, the confidence intervals by the second method could be rather conservative in this case.

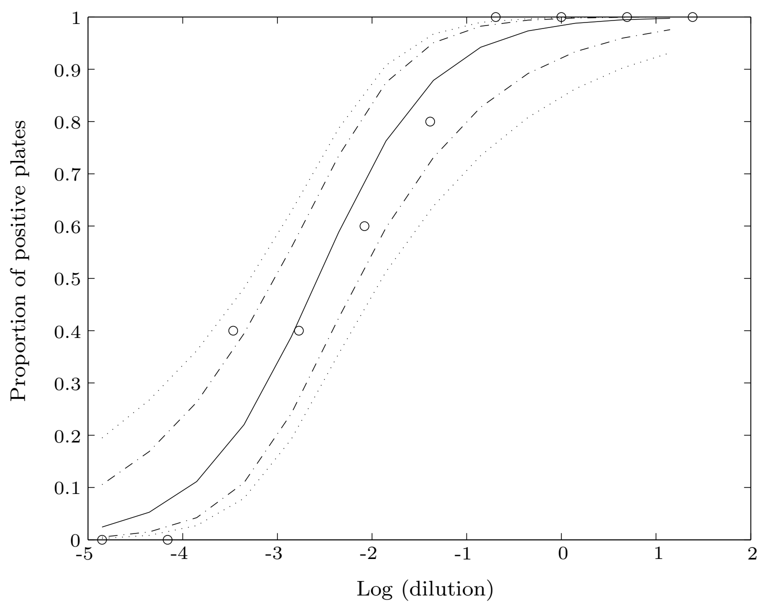

Figure 3.6. Comparison of the two-sided (pointwise) confidence intervals constructed from the naive method by l_c (*dotted* lines) and the proposed method by l_c^* (*dot-dash* lines). The solid line represents the pointwise MLE of the proportions of positive plates.

Appendix.

Proof of Proposition 2.1 (i) Let D be the differential operator. It is known that the Hermite polynomials $H_j(x), j = 0, 1, 2, \ldots$, defined by $-H_j(x)\phi(x) = (-\mathrm{D})^{j+1}\Phi(x)$, form an orthogonal basis with respect to the weight function $\phi(x)$. Then there exist polynomials $r_1(x)$ and $r_2(x)$, having a factor x, such that $p_j(x)\phi(x) = r_j(-\mathrm{D})\Phi(x)$, $j = 1, 2$. Hence the characteristic function of W can be written as

$$\begin{aligned}C_W(t) &= \int e^{itx}\mathrm{d}P(W \le x)\\ &= \int e^{itx}\mathrm{d}\Phi(x) + n^{-\frac{1}{2}}\int e^{itx}\mathrm{d}\{r_1(-\mathrm{D})\Phi(x)\} + n^{-1}\int e^{itx}\mathrm{d}\{r_2(-\mathrm{D})\Phi(x)\} + \mathrm{o}(n^{-1}).\end{aligned}$$

Observing that $\int e^{itx}\mathrm{d}\{r_j(-\mathrm{D})\Phi(x)\} = r_j(it)e^{-t^2/2}$, we find

$$C_W(t) = e^{-\frac{t^2}{2}}\big[1 + n^{-\frac{1}{2}}r_1(it) + n^{-1}r_2(it)\big] + \mathrm{o}(n^{-1}). \tag{A.1}$$

On the other hand, assume that there are polynomials $q_1(\cdot)$ and $q_2(\cdot)$ such that $W \overset{d}{=} Z + n^{-1/2}q_1(Z) + n^{-1}q_2(Z)$. Then we can write

$$\begin{aligned} C_W(t) = &\int e^{itz}\mathrm{d}\Phi(z) + \frac{1}{\sqrt{n}}\int e^{itz}it\cdot q_1(z)\phi(z)\mathrm{d}z \\ &+\frac{1}{n}\int e^{itz}\big[it\cdot q_2(z) + \frac{(it)^2}{2}q_1^2(z)\big]\phi(z)\mathrm{d}z + \mathrm{o}(n^{-1}). \end{aligned} \tag{A.2}$$

Likewise, there exist polynomials $r_j^*(x)$, $j = 1,2,3$, having a factor x such that $q_j(z)\phi(z) = r_j^*(-\mathrm{D})\Phi(z)$, $j = 1,2$, $q_1^2(z)\phi(z) = r_3^*(-\mathrm{D})\Phi(z)$. Note that for such polynomials r_j^*, integration by parts yields

$$\int e^{itz}it\, r_j^*(-\mathrm{D})\Phi(z)\mathrm{d}z = -\int e^{itz}\mathrm{d}\{r_j^*(-\mathrm{D})\Phi(z)\}.$$

Therefore the RHS of (A.2) equals

$$\begin{aligned} &e^{-\frac{t^2}{2}} + \frac{1}{\sqrt{n}}\big[-\int e^{itz}\mathrm{d}\{r_1^*(-\mathrm{D})\Phi(z)\}\big] \\ &\quad+\frac{1}{n}\big[-\int e^{itz}\mathrm{d}\{r_2^*(-\mathrm{D})\Phi(z)\} - \frac{it}{2}\int e^{itz}\mathrm{d}\{r_3^*(-\mathrm{D})\Phi(z)\}\big] + \mathrm{o}(n^{-1}) \\ &= e^{-\frac{t^2}{2}}\big\{1 - n^{-\frac{1}{2}}r_1^*(it) - n^{-1}\big[r_2^*(it) + \frac{it}{2}r_3^*(it)\big]\big\} + \mathrm{o}(n^{-1}). \end{aligned} \tag{A.3}$$

Comparing the coefficients of $n^{-1/2}$ and n^{-1} in (A.1) and (A.3), we get

$$\begin{cases} r_1(x) = -r_1^*(x), \\ r_2(x) = -r_2^*(x) - \frac{x}{2}r_3^*(x). \end{cases} \tag{A.4}$$

Then

$$\begin{aligned} p_1(x)\phi(x) &= r_1(-\mathrm{D})\Phi(x) = -r_1^*(-\mathrm{D})\Phi(x) = -q_1(x)\phi(x), \\ p_2(x)\phi(x) &= r_2(-\mathrm{D})\Phi(x) = \big[-r_2^*(-\mathrm{D}) + \frac{\mathrm{D}}{2}r_3^*(-\mathrm{D})\big]\Phi(x) \\ &= \big[-q_2(x) + q_1(x)q_1'(x) - \frac{x}{2}q_1^2(x)\big]\phi(x). \end{aligned} \tag{A.5}$$

Finally (2.9) follows, after solving for $q_1(x)$ and $q_2(x)$ by $p_1(x)$ and $p_2(x)$ in (A.5).

(ii) To show (2.10), simply substitute W from the RHS of (2.8), and use a Taylor expansion. We get

$$\begin{aligned} &W + n^{-\frac{1}{2}}h_1(W) + n^{-1}h_2(W) \\ &\overset{d}{=} Z + n^{-\frac{1}{2}}[q_1(Z) + h_1(Z)] + n^{-1}[q_2(Z) + h_1'(Z)q_1(Z) + h_2(Z)]. \end{aligned}$$

Then (2.11) immediately implies (2.10).

In fact it is easy to verify that (2.8) and (2.10) are equivalent if and only if (2.11) holds.

(iii) From the additional assumption, we first have $P(|W| \le x) = (2\Phi(x) - 1) + 2n^{-1}p_2(x)\phi(x) + \mathrm{o}(n^{-1})$. Then along the same argument as in the proof of (i) we obtain the desired representations. The details are omitted.

Acknowledgement

The authors are thankful to the Editor and one referee for their helpful comments on the initial manuscript. The first author has been supported by a dissertation fellowship at Bowling Green State University.

References

Abramovitch, L. and Singh, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *Ann. Statist.* **13**, 116-132.

Bartlett, M. S. (1953). Approximate confidence intervals. *Biometrika* **40**, 12-19.

Bhattacharya, R. N. and Rao, R. R. (1976). *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

Brown, L. D., Cai, T. T. and DasGupta, A. (2002). Confidence intervals for a binomial proportion and Edgeworth expansions. *Ann. Statist.* **30**, 160-201.

Brown, L. D., Cai, T. T. and DasGupta, A. (2003). Interval estimation in exponential families. *Statist. Sinica* **13**, 19-49.

Fisher, R. A. and Yates, F. (1963). *Statistical Tables For Biological, Agricultural and Medical Research, 6th Ed.* Hafner, New York.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansions.* Springer-Verlag, New York.

Johnson, N. J. (1978). Modified t tests and confidence intervals for asymmetrical populations. *J. Amer. Statist. Assoc.* **73**, 536-544.

Konishi, S. (1991). Normalizing transformation and bootstrap confidence intervals. *Ann. Statist.* **19**, 2209-2225.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models.* 2nd edition. Chapman & Hall.

McCulloch, C. E. and Searle, S. R. (2001) *Generalized, Linear and Mixed Models.* Wiley, New York.

Sun, J., Loader, C. and McCormick, W. P. (2000). Confidence bands in generalized linear models. *Ann. Statist.* **28**, 429-460.

Xu, J. and Gupta, A. K. (2003). Asymptotic expansions of the distributions of some test statistics in generalized linear models. *Department of Mathematics and Statistics, Bowling Green State University, Technical Report* No. 03-15.

Xu, J. and Gupta, A. K. (2004). Inproved confidence regions for a mean vector under general condition. *Department of Mathematics and Statistics, Bowling Green State University, Technical Report* No. 04-02.

Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43402, U.S.A.

E-mail: jinx@bgnet.bgsu.edu

Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43402, U.S.A.

E-mail: gupta@bgnet.bgsu.edu

(Received December 2003; accepted October 2004)