The Annals of Statistics
2000, Vol. 28, No. 2, 429–460

CONFIDENCE BANDS IN GENERALIZED LINEAR MODELS

BY JIAYANG SUN,[1] CATHERINE LOADER AND WILLIAM P. MCCORMICK

Case Western Reserve University, Bell Laboratories and University of Georgia

Generalized linear models (GLM) include many useful models. This paper studies simultaneous confidence regions for the mean response function in these models. The coverage probabilities of these regions are related to tail probabilities of maxima of Gaussian random fields, asymptotically, and hence, the so-called tube formula is applicable without any modification. However, in the generalized linear models, the errors are often non-additive and non-Gaussian and may be discrete. This poses a challenge to the accuracy of the approximation by the tube formula in the moderate sample situation. Here two alternative approaches are considered. These approaches are based on an Edgeworth expansion for the distribution of a maximum likelihood estimator and a version of Skorohod's representation theorem, which are used to convert an error term (which is of order $n^{-1/2}$ in one-sided confidence regions and of n^{-1} in two-sided confidence regions) from the Edgeworth expansion to a "bias" term. The bias is then estimated and corrected in two ways to adjust the approximation formula. Examples and simulations show that our methods are viable and complementary to existing methods. An application to insect data is provided. Code for implementing our procedures is available via the software parfit.

1. Introduction. Generalized linear models (GLM) include many useful models, for example, linear regression models for normal response data, linear logistic models for binary response data and log-linear models for categorical response data. It is important to quantify the shape, presence of some special features and uncertainty of the estimates of the mean response curve or the regression function. For this purpose, this article studies informative simultaneous confidence regions (SCR hereafter) for the mean response curve under generalized linear models.

The Scheffé (1959) method may be used to construct a SCR for a linear parametric regression function over the *whole* predictor space when errors are i.i.d. normal and additive. In more practical cases when the domain of interest is a subset of the whole predictor space, Naiman (1987, 1990) and Sun and Loader (1994) provided SCRs for a regression function over the subset using the "tube method" [see also Knowles and Siegmund (1989), Johansen and Johnstone (1990)]. These regions are robust for additive models with contaminated normal errors and errors from spherical symmetric distributions

Received October 1995; revised July 1999.

[1]Supported in part by NSF Grants DMS-95-04515, DMS-96-26108 and DM-98-70544.

AMS 1991 *subject classifications*. Primary 62F25, 62J12; secondary 60G70, 60G15, 62G07, 62E20.

Key words and phrases. Tube formula, Edgeworth expansion, maximum of Gaussian random fields, regression, simultaneous confidence bands.

[cf. Loader and Sun (1997)]. The "tube formula" in principle also works without modification for models with any nonnormal additive independent errors when the sample size is large as the central limit theorem applies. In other cases, this tube formula without any modification (called naive SCR hereafter) may fail, as shown in Loader and Sun (1997). It is encouraging that the tube formula can be modified and combined with other techniques to extend to the cases with independent heteroscedastic or correlated additive errors, after an adjustment to account for the extra variation introduced by estimation of unknown weights in a weighted least squares regression or the structured covariance in the correlated case [cf. Faraway and Sun (1995) and Sun, Raz and Faraway (1999)]. However, in the generalized linear model, the errors are often nonadditive (unless they are normal) and (response) variables are discrete. This poses a challenge to the tube method used in the above work when the sample size is moderate.

Here two remedies are proposed. They apply the Edgeworth expansion in the middle of approximations rather than the normal approximation for a normalized process and then use the idea of the Skorohod construction to convert the first order error term from the Edgeworth expansion (in terms of distribution) to a bias term (in terms of the difference of the process and its limit, in a suitable sample space). The converted expansion may be called an *inverse Edgeworth expansion*. Our corrections then proceed in two ways in "correcting" the bias term from the inverse Edgeworth expansion.

More details are as follows. In Section 2, the GLM and some basic SCRs are described. In Section 3, several inverse Edgeworth expansions are developed. The first expansion is called basic inverse Edgeworth expansion in this paper. The inverse Edgeworth expansion is of interest in itself and can be considered as a generalization of the Cornish–Fisher expansion for a partial sum of i.i.d. random variables. Other inverse Edgeworth expansions are generalized from the basic inverse Edgeworth expansion for some special processes to make corrections in Section 4. In Section 4, a first-level approximation to the coverage probabilities of the basic SCRs is provided; corrections of the basic SCRs are derived in two ways for each of two types of SCRs: one- and two-sided SCRs. In the case of the one-sided SCR, the error term in the Edgeworth expansion is at the rate of $n^{-1/2}$ and depends on the bias, skewness and variance of a normalized process; in the case of two-sided SCRs, the error term is of order n^{-1} and depends on the kurtosis and a secondary effect of the skewness of the process. Thus, the corrections for one-sided and two-sided SCRs are different. Our first way of correction is to apply the tube formula to some modified processes. Our second way of correction uses the method of bias correction in Sun and Loader (1994), which is equivalent to shrinking and expanding the width of the basic SCRs, based on a bound of the bias term. Examples and simulations are shown in Section 5. Concluding remarks and comparisons with alternative methods [e.g., Knafl, Sacks and Ylvisaker (1985), Hall and Titterington (1988) and Härdle and Marron (1991)] are summarized in Section 6. Proofs and other examples can be found in the Appendix. Those who wish to use our SCRs right away may read Table 3 in Section 5 directly, though reading the description of GLM in Section 2 and the concluding remarks in Section 6 is always helpful. Those who wish to check the meaning of a notation may consult the list of notations at the end of this paper, arranged roughly in the order of their appearances in the paper.

2. Model and basic simultaneous confidence regions.

Model. Consider the usual generalized linear model (GLM) specified by a *link function* and a *random structure*. Here the (anti) link function g links n independent observable pairs of predictor x_i and response variable Y_i by

$$-\infty < E(Y_i|x_i) = g(\eta(x_i)) < \infty \quad \text{for } i = 1\dots n,$$

where $E(Y|x)$ is the conditional expectation of a response Y given a predictor x, $\eta(x) = z(x)^T\beta$ is a linear predictor, β is a q-dimensional unknown parameter and z is a q-dimensional known covariate, for example, $z(x) = (1, x, x^2)$. The *random structure* assumes that the density f of Y_i belongs to an exponential family,

$$f(y;\theta_i) = \exp\{y\theta_i - b(\theta_i) + a(y)\} \tag{1}$$

with respect to some σ-finite measure $d\Lambda(y)$, for some known functions b and a and a natural parameter $\theta_i = \theta(x_i)$. The natural parameter is a function of the linear predictor $\eta(x_i)$. Our problem of interest is to find under the GLM an informative simultaneous confidence region for the mean response function $E(Y|x)$ as a function of x over a domain of interest $\mathscr{X} \subseteq \mathbb{R}^d$ for some $d > 0$. In this paper, for simplicity we only study the case $d = 1$. The extension to a high-dimensional d is possible since both the tube formula and the inverse Edgeworth expansion (can) have their multivariate versions.

Throughout this paper, a subscripted function denotes the function evaluated at x_i, for example $z_i = z(x_i)$ and $\theta_i = \theta(x_i)$, and b' denotes the first derivative of b, b'' the second derivative and $b^{(3)}$ the third derivative, etc.

Basic simultaneous confidence regions. In most cases, the link function g is chosen to be the canonical link. The canonical link requires that the natural parameter $\theta = \theta(x) = \eta(x)$, so $g(\eta) = g(\theta) = E(Y|x) = b'(\theta)$, which is a known monotone function of η since $b''(\theta) = \mathrm{var}(Y|x) > 0$. Hence finding a SCR for the mean response function $E(Y|x)$ is equivalent to finding a SCR for the linear predictor $\eta(x)$, for $x \in \mathscr{X}$.

A basic two-sided SCR for $\eta(x)$ is one centered around its estimator $\hat\eta(x) = z(x)^T\hat\beta$,

$$l(x) \equiv l_c(x) = (\hat\eta(x) - c\hat\sigma(x), \hat\eta(x) + c\hat\sigma(x)) \qquad \forall x \in \mathscr{X} \tag{2}$$

for some $c > 0$, where $\hat\beta = \hat\beta_n$ is the maximum likelihood estimator of β, and $\hat\sigma^2(x)$ is the estimated asymptotic variance of $\hat\eta(x)$. Equations (6) and (7) below have their detailed expressions. For a prescribed level $0 < 1 - \alpha < 1$, c is such that

$$1 - \alpha = \mathrm{pr}\{\eta(x) \in l_c(x),\ \forall\, x \in \mathscr{X}\}. \tag{3}$$

Once c is found, we have a SCR $l_c(x)$ for $\eta(x)$ and hence a SCR $g(l_c(x))$ for the mean response function $E(Y|x)$ for $x \in \mathscr{X}$. Note that for an approximate pointwise CI at a fixed x, c is simply $\Phi^{-1}(1-\alpha/2)$, the upper $\alpha/2$ quantile of the standard normal distribution.

Sometimes one-sided confidence bounds are of interest. For example, in quality control applications, one might want upper bounds on the failure rates of components under various operating conditions. Our one-sided basic upper and lower SCRs for $\eta(x)$ are of the form

$$\begin{aligned} l^u(x) &= (-\infty, \hat{\eta}(x) + c^u\hat{\sigma}(x)) \qquad \forall\, x \in \mathscr{X}, \\ l^l(x) &= (\hat{\eta}(x) - c^l\hat{\sigma}(x), \infty) \qquad \forall\, x \in \mathscr{X}, \end{aligned} \tag{4}$$

respectively, where c^u and c^l are such that

$$\mathrm{pr}\{\eta(x) \in l^u(x),\ \forall\, x \in \mathscr{X}\} = 1 - \alpha = \mathrm{pr}\{\eta(x) \in l^l(x),\ \forall\, x \in \mathscr{X}\}. \tag{5}$$

We shall see in Section 4 that $c^l = c^u$, using a first-level approximation to (5).

Clearly, the estimator $\hat{\beta}$ is a solution of the MLE equation

$$\sum_{i=1}^{n} z_i[Y_i - b'(\theta_i)] = 0. \tag{6}$$

The asymptotic variance $\sigma^2(x)$ of $\hat{\eta}(x)$ and its estimator are

$$\sigma^2(x) = z(x)^T I_n^{-1}(\beta) z(x), \qquad \hat{\sigma}^2(x) = z(x)^T I_n^{-1}(\hat{\beta}) z(x), \tag{7}$$

where $\sigma(x)/\hat{\sigma}(x) = 1 + O_p(n^{-1/2})$ [obviously, under the Cramér conditions that lead to $\hat{\beta} - \beta = O_p(n^{-1/2})$] and $I_n(\beta)$ is the $q \times q$ Fisher information matrix about β in the sample (x_i, Y_i), $i = 1, \ldots, n$. This Fisher information can be expressed as

$$I_n(\beta) = \sum_{i=1}^{n} b''(\theta_i) z_i z_i^T = X^T \Sigma X, \tag{8}$$

where $z_i = z(x_i)$, $X^T = (z_1, \ldots, z_n)$ is the $n \times q$ design matrix and

$$\Sigma = \mathrm{diag}(\mathrm{var}(Y_1), \ldots, \mathrm{var}(Y_n)) = \mathrm{diag}(b''(\theta_1), \ldots, b''(\theta_n))$$

is the variance matrix of the responses. Of course, $\hat{\beta}$ may be a biased estimator of β. We will not merely recenter or shift the region in (2) based on an estimator of the bias, as this may increase the variance of the adjusted estimator and sometimes is worse than no correction [cf. Loader (1993) and Sun and Loader (1994)]. Instead, to achieve a prescribed level, we shall use two methods of bias corrections discussed at the end of Section 1.

Discussion. The following development and corrections of basic SCRs are for the case that the link is canonical. However, it is apparent that our method also applies to the case with known noncanonical link. A discussion for the case of an unknown link function is given in Section 6.

3. Inverse Edgeworth expansions. In this section, inverse Edgeworth expansions are developed using Skorohod construction. Examples of relevant coefficients in the inverse Edgeworth expansions, such as bias, skewness and lattice term, are presented in the Appendix. For two sequences of distributions, F_n and G_n, it will be convenient to have a notation to denote equality of distributions up to $o(n^{-1/2})$ or $o(n^{-1})$. We write $F_n =^d G_n$ to indicate that $F_n(x) - G_n(x) = o(n^{-1/2})$ or $o(n^{-1})$, uniformly in x. Whether it is $o(n^{-1/2})$ or $o(n^{-1})$ should be clear from the last term in F_n or G_n. As is customary, we also write $X_n =^d Y_n$ to indicate that the corresponding sequence of distributions stand in the relation of equality up to $o(n^{-1/2})$ or $o(n^{-1})$ and we write $\langle x, y\rangle$ to denote the inner product between two arguments x and y.

Inverse Edgeworth expansions of vectors. The heuristics for developing an inverse Edgeworth expansion is as follows. When β is one-dimensional, for a normalizing constant a_n, following Skovgaard (1981a, b) we have an Edgeworth expansion for the random sequence $a_n(\hat{\beta}_n - \beta)$ given by

$$\mathrm{pr}\{a_n(\hat{\beta}_n - \beta) \le t\} = \Phi(t) + R_n(t) \equiv F_n(t),$$

where $R_n(t) = O(n^{-1/2})$ depends on the bias and skewness of $a_n(\hat{\beta}_n - \beta)$ and $\Phi(t)$ is the standard normal cumulative distribution function. (When β is $q > 1$-dimensional, a similar expansion exists for $\langle v, a_n(\hat{\beta}_n - \beta)\rangle$ for any $v \in \mathbb{R}^q$.) Thus, for some $Q_n(t) = O(n^{-1/2})$,

$$a_n(\hat{\beta}_n - \beta) \stackrel{d}{=} F_n^{-1}(U) = \Phi^{-1}(U) + Q_n(U) \stackrel{d}{=} Z + Q_n(\Phi(Z)), \tag{9}$$

where U is a uniform random variable on (0,1) and Z is a standard normal random variable. In this way, the $O(n^{-1/2})$ error term R_n in the convergence of the distribution is passed onto a $O_p(n^{-1/2})$ "bias" term $Q_n(\Phi(Z))$ between the random sequence and its limit, in a suitable sample space. This is similar to that of Skorohod construction. A detailed expression of Q_n is given in (10) of the following proposition.

PROPOSITION 3.1 (Basic inverse Edgeworth expansion). *Let $\underline{\lambda}_n \le \overline{\lambda}_n$ denote the smallest and largest eigenvalues of A_n where $A_n(\beta) \equiv I_n(\beta)/n$ – a rescaled Fisher information, and let $B_n = B_n(\beta)$ be the upper* Cholesky triangle *in the decomposition $B_n^T B_n = A_n$. Suppose that $E|Y_i|^3 < \infty$ for all i and $0 < \liminf \underline{\lambda}_n \le \limsup \overline{\lambda}_n < \infty$. Then for any $v \in S^{q-1}$ (the q-dimensional unit sphere) and a random variable $Z \sim N(0,1)$, we have the following basic inverse Edgeworth expansion:*

$$\big\langle v, \sqrt{n}B_n(\hat{\beta}_n - \beta)\big\rangle \stackrel{d}{=} Z + \tfrac{1}{6}\tau_n(v)\big[Z^2 - 1\big] + \langle v, \mu_n\rangle + T_n(v) \tag{10}$$

as $n \to \infty$. Here $\langle v, \mu_n\rangle$, $\tau_n(v)$ and $T_n(v)$ are the bias, skewness and lattice terms of $\big\langle v, \sqrt{n}B_n(\hat{\beta}_n - \beta)\big\rangle$ given in (13), (14) *and* (15) *below.*

The key formula for deriving the bias, skewness and lattice term in (10) is the following asymptotic expression of $B_n\sqrt{n}(\hat{\beta}_n - \beta)$:

$$B_n\sqrt{n}(\hat{\beta}_n - \beta) = \xi - \frac{1}{2n^{3/2}} \sum_{i=1}^{n} b^{(3)}(\theta_i)\langle u_i, \xi\rangle^2 u_i + o_p\left(\frac{1}{n^{1/2}}\right), \tag{11}$$

which comes from a second-order expansion of the MLE equation (6). Here $u_i = \left(B_n^T\right)^{-1} z_i$ is a normalized covariate of z_i, and

$$\xi = \xi(n) = \frac{1}{n^{1/2}} \sum_{i=1}^{n} u_i\left[Y_i - b'(\theta_i)\right] \tag{12}$$

is the normalized random variable such that $E(\xi) = 0$ and $E(\xi\xi^T) = I_q$, the q-dimensional identity matrix. The moment relationship about ξ here follows easily from (8). See the Appendix for its detailed derivation and the proof of Proposition 3.1.

Based on the asymptotic expansion (11), one can simply compute the first and third moments of $B_n\sqrt{n}(\hat{\beta}_n - \beta)$ to get the bias $\langle v, \mu_n\rangle$ and skewness $\tau_n(v)$,

$$\mu = \mu_n = E\left[B_n\sqrt{n}(\hat{\beta}_n - \beta)\right] = \frac{-1}{2n^{3/2}} \sum_{i=1}^{n} b^{(3)}(\theta_i)\|u_i\|^2 u_i, \tag{13}$$

$$\tau_n(v) = E\left\langle v, B_n\sqrt{n}(\hat{\beta}_n - \beta) - \mu\right\rangle^3 = \frac{-2}{n^{3/2}} \sum_{i=1}^{n} b^{(3)}(\theta_i)\langle u_i, v\rangle^3, \tag{14}$$

up to $o(n^{-1/2})$, for any $v \in S^{q-1}$; compare (A.3).

The bias and skewness terms are of $O(n^{-1/2})$. If the distribution of $\langle v, \xi\rangle$ is lattice, clearly there is an additional $n^{-1/2}$ order (lattice) term $T_n(v)$ in (10),

$$T_n(v) = \begin{cases} \dfrac{h(v)}{n^{1/2}} H\left(\dfrac{n^{1/2}}{h(v)} Z\right), & \text{if } \langle \xi, v\rangle \text{ has a lattice distribution,} \\ 0, & \text{if } \langle \xi, v\rangle \text{ does not have a lattice distribution,} \end{cases} \tag{15}$$

where $h(v)$ is the span of $\langle v, \xi\rangle$, and $H(x)$ is a periodic function with a period 1 and such that $H(x) = x - 1/2$ for $0 \le x \le 1$.

It is easy to check when the lattice term is zero. A necessary condition for $\langle v, \xi\rangle$ to have a lattice distribution is that at least one component of $\xi^T = (\xi_1, \dots, \xi_q)$ has a lattice distribution. In order words, there are constants a_i and $h_i > 0$ for some i such that

$$\operatorname{pr}(\xi_i = a_i + h_i k \text{ for some } k = \dots, -2, -1, 0, 1, 2, \dots) = 1.$$

The largest real number h_i such that the above statement holds is called the span. Based on this necessary condition, it is rare to need a nonzero lattice term even if the response variable is discrete. See the examples in the Appendix.

In contrast to the one-sided case, in the case of a two-sided SCR the main effect of skewness is of $O(n^{-1})$ instead of $O(n^{-1/2})$, because the $O(n^{-1/2})$ terms cancel at two boundaries. Indeed, as in Hall [(1992), page 49],

$$\mathrm{pr}\{|a_n(\hat{\beta}_n - \beta)| \le t\} = 2\Phi(t) - 1 + R_n'(t) \equiv F_n(t),$$

where $R_n'(t) = O(n^{-1})$ depends on the kurtosis and a secondary effect of the skewness and variance of $a_n(\hat{\beta}_n - \beta)$. Then, similar to the idea for obtaining (9), one obtains

$$|a_n(\hat{\beta}_n - \beta)| \stackrel{d}{=} |Z| + Q_n'(\Phi(Z)). \tag{16}$$

The p_2 in (25) below is an example of Q_n'.

Inverse Edgeworth expansions of processes. Let $W_n(x)$ and $W_n^*(x)$ denote two normalized processes

$$W_n(x) := \frac{\hat{\eta}(x) - \eta(x)}{\hat{\sigma}(x)} \quad \text{and} \quad W_n^*(x) := \frac{\hat{\eta}(x) - \eta(x)}{\sigma(x)}. \tag{17}$$

It is straightforward to show that under some regularity conditions [e.g., the exponential family specified by (1) is minimal and steep in the sense of Brown (1986), and the natural parameter space contains at least one interior point], the family $\{W_n(x)\}$ is tight and

$$W_n(x) \xrightarrow{d} W(x) \tag{18}$$

and hence $W_n^*(x) \longrightarrow^d W(x)$, as $n \to \infty$. Here $W(x)$ is a standard Gaussian random field with mean zero and covariance function $\rho(x, x') = \langle s(x'), s(x)\rangle$, where $\rho(x, x') = 1$ if $x = x'$, $s(x) = \lim_{n\to\infty} s_n(x)$ and

$$s_n(x) = \frac{(B_n^T)^{-1} z(x)}{\sqrt{n}\sigma(x)} = \frac{(B_n^T)^{-1} z(x)}{\sqrt{z^T A_n^{-1} z}}. \tag{19}$$

Since $\hat{\eta}(x) = z(x)^T \hat{\beta}$ and $W_n^*(x) = \langle s_n(x), B_n\sqrt{n}(\hat{\beta} - \beta)\rangle$, it follows from (10) and similar steps in establishing (18) that an *inverse Edgeworth expansion for* $W_n^*(x)$ is

$$W_n^*(x) \stackrel{d}{=} W(x) + \tfrac{1}{6}\tau_n(s_n(x))[W(x)^2 - 1] + \langle s_n(x), \mu_n\rangle + T_n(s_n(x)), \tag{20}$$

under the same regularity conditions of Proposition 3.1.

However, the inverse Edgeworth expansion for $W_n(x)$ differs from (20), since $W_n(x) = W_n^*(x)\sigma(x)/\hat{\sigma}(x)$ and $\sigma(x)/\hat{\sigma}(x) = 1 + O_p(n^{-1/2})$. Indeed, multiplying W_n^* [obtained by taking the inner product of the expression in (11) with $s_n = s_n(x)$] by an expansion of $\sigma(x)/\hat{\sigma}(x)$, we have up to $o_p(n^{-1/2})$,

$$\begin{aligned} W_n(x) = \langle s_n, \xi\rangle + \frac{1}{2n^{3/2}} \sum_{i=1}^n b^{(3)}(\theta_i) \\ \times\big[\langle s_n, u_i\rangle^2 \langle u_i, \xi\rangle\langle s_n, \xi\rangle - \langle u_i, s_n\rangle\langle u_i, \xi\rangle^2\big]. \end{aligned} \tag{21}$$

It then follows by straightforward calculations that the mean, covariance and skewness of $W_n(x)$, up to $o(n^{-1/2})$, are

$$\mu'_n(x) := E(W_n(x)) = \frac{1}{2n^{3/2}} \sum_{i=1}^{n} b^{(3)}(\theta_i)\big[\langle s(x), u_i\rangle^3 - \langle s(x), u_i\rangle \|u_i\|^2\big],$$

$$\operatorname{cov}(W_n(x), W_n(x')) = \langle s(x), s(x')\rangle, \tag{22}$$

$$\tau'_n(x) := \operatorname{skew}(W_n(x)) = \frac{1}{n^{3/2}} \sum_{i=1}^{n} b^{(3)}(\theta_i)\langle u_i,\ s(x)\rangle^3.$$

See the Appendix for a more detailed derivation of (21) and (22). Consequently, an inverse Edgeworth expansion of $W_n(x)$ is

$$W_n(x) \stackrel{d}{=} W(x) + \tfrac{1}{6}\tau'_n(x)[W(x)^2 - 1] + \mu'_n(x) + T_n(s(x)). \tag{23}$$

Recall from (3) and (5) that our goal is to choose $c > 0$, $c^u > 0$ and $c^l > 0$ such that $\alpha = P(c)$ and $P^u(c^u) = \alpha = P^l(c^l)$, where

$$\begin{aligned} P(c) &= \operatorname{pr}\Big\{\sup_{x\in\mathscr{X}} |W_n(x)| > c\Big\},\\ P^u(c^u) &= \operatorname{pr}\Big\{\inf_{x\in\mathscr{X}} W_n(x) < -c^u\Big\},\\ P^l(c^l) &= \operatorname{pr}\Big\{\sup_{x\in\mathscr{X}} W_n(x) > c^l\Big\}. \end{aligned} \tag{24}$$

Clearly, W_n has a mean and a skewness much closer to zero than those of W_n^*, and the skewness of W_n and W_n^* have different signs; see (13), (14) and (22). Thus, an application of (20) (intended for W_n^*) for correcting c^u in l^u (related to W_n) would result in *overcorrections*. In Section 4, we shall use (23) to build corrected one-sided SCRs.

Tracing the same arguments for (23) and using (16), we have a two-sided inverse Edgeworth expansion for $W_n(x)$,

$$|W_n(x)| \stackrel{d}{=} |W(x)| - p_2(x, W_n(x)), \tag{25}$$

where the meaning of p_2 is similar to that in Hall [(1992), pages 48–50] so that

$$p_2(x, z) = -z\bigg\{\frac{1}{2}\big[\kappa_2(x) - 1 + \kappa_1^2(x)\big] \tag{26}$$

$$+ \frac{1}{24}\big[\kappa_4(x) + 4\kappa_1(x)\kappa_3(x)\big](z^2 - 3)$$

$$+ \frac{1}{72}\kappa_3^2(x)(z^4 - 10z^2 + 15)\bigg\} = O\Big(\frac{1}{n}\Big). \tag{27}$$

Here κ_i's are the moments of $W_n(x)$, up to $o(n^{-1})$,

$$
\begin{aligned}
&\kappa_1(x) = EW_n(x) = \mu_n'(x), \qquad \kappa_3(x) = E[W_n(x) - EW_n(x)]^3 = \tau_n'(x),\\
&\kappa_2(x) = E[W_n(x) - EW_n(x)]^2 = 1 + \tfrac{1}{2}C_1 - \tfrac{1}{2}C_2\\
&\qquad\qquad -3C_3 + \tfrac{1}{2}C_4 + C_6 - \tfrac{1}{2}C_7 + \tfrac{7}{4}C_8,\\
&\kappa_4(x) = E[W_n(x) - EW_n(x)]^4 - 3\kappa_2^2(x)\\
&\qquad\quad = -9C_3 + 3C_6 + 6C_8 + 3C_9
\end{aligned}
\tag{28}
$$

with $u_i = (B_n^T)^{-1} z_i$ and

$$
\begin{aligned}
C_1 &= \frac{1}{n^3}\sum_i\sum_j b_i^{(3)} b_j^{(3)} \langle s(x),\, u_i\rangle \langle s(x),\, u_j\rangle \langle u_i,\, u_j\rangle^2 = C_5,\\
C_2 &= \frac{1}{n^2}\sum_i b_i^{(4)} \langle s(x),\, u_i\rangle^2 \langle u_i,\, u_i\rangle,\\
C_3 &= \frac{1}{n^3}\sum_i\sum_j b_i^{(3)} b_j^{(3)} \langle s(x),\, u_i\rangle^2 \langle s(x),\, u_j\rangle^2 \langle u_i,\, u_j\rangle,\\
C_4 &= \frac{1}{n^3}\sum_i\sum_j b_i^{(3)} b_j^{(3)} \langle s(x),\, u_i\rangle^2 \langle u_i,\, u_j\rangle \langle u_j,\, u_j\rangle,\\
C_6 &= \frac{1}{n^2}\sum_i b_i^{(4)} \langle s(x), u_i\rangle^4,\\
C_7 &= \frac{1}{n^3}\sum_i\sum_j b_i^{(3)} b_j^{(3)} \langle s(x),\, u_i\rangle \langle s(x),\, u_j\rangle^3 \langle u_i,\, u_i\rangle,\\
C_8 &= \frac{1}{n^3}\sum_i\sum_j b_i^{(3)} b_j^{(3)} \langle s(x),\, u_i\rangle^3 \langle s(x),\, u_j\rangle^3 = [\kappa_3(x)]^2,\\
C_9 &= \frac{1}{n^2}\sum_i \big[b_i^{(2)}\big]^2 \langle s(x),\, u_i\rangle^4.
\end{aligned}
\tag{29}
$$

Similarly, let $W_n^{(0)}(x)$ be a normalized process of $W_n(x)$,

$$
W_n^{(0)}(x) = (W_n(x) - \hat{\kappa}_1(x)) \Big/ \sqrt{\hat{\kappa}_2(x)}; \tag{30}
$$

then a two-sided inverse Edgeworth expansion of $W_n^{(0)}(x)$ is

$$
\left|W_n^{(0)}(x)\right| \stackrel{d}{=} |W(x)| - q_2(x,\, W_n^{(0)}(x)), \tag{31}
$$

where q_2 is the counterpart of p_2 for $W_n^{(0)}(x)$,

$$
\begin{aligned}
q_2(x,\, z) = -z\big\{&\tfrac{1}{2}\kappa_{2,\,2}^s(x) + \tfrac{1}{24}\kappa_{4,\,1}^s(x)(z^2 - 3)\\
&+ \tfrac{1}{72}\big[\kappa_{3,\,1}^s(x)\big]^2 (z^4 - 10z^2 + 15)\big\},
\end{aligned}
\tag{32}
$$

$$\begin{aligned} \kappa^s_{2,2}(x) &= C_3 - \tfrac{3}{2}C_8 - C_5 - C_4 + \tfrac{1}{2}C_7 + C_6 - C_2, \\ \kappa^s_{3,1}(x) &= \kappa_3(x) = \tau'_n(x), \\ \kappa^s_{4,1}(x) &= -3C_3 - 6C_4 - 6C_5 + 3C_6 + 3C_7 - 3C_8 + 3C_9. \end{aligned} \tag{33}$$

Here, $q_2(x, z) = O(n^{-1}) = p_2(x, z)$. From their expressions, it is obvious that the computation involving the n^{-1} correction term increases drastically from that of $n^{-1/2}$. Nevertheless, these formulas are easy to compute in a computer; a user may use our software `parfit` to automate the computation.

4. Approximations and corrections. In this section, a first-level approximation is developed for the coverage probabilities of basic SCRs in (2) and (4). The inverse Edgeworth expansions in Section 3 are used to derive corrected SCRs, in two different ways.

Approximations. In the typical case that $s(x)$ does not self-overlap [in most cases it does not; cf. Sun and Loader (1994)], it is straightforward to apply the weak convergence (18), tube formula in Sun (1993) and a boundary correction [Sun and Loader (1994)] to get

$$P^l(c) \approx \mathrm{pr}\left(\sup_{x\in\mathscr{X}} W(x) > c\right) \approx \frac{\kappa_0}{2\pi}\exp(-c^2/2) + \delta\big(1-\Phi(c)\big)$$

where $P^l(c)$ is in (24), κ_0 is the length of $\{s(x): x \in \mathscr{X}\}$,

$$\kappa_0 = \int_{x\in\mathscr{X}} \|s'(x)\|\, dx, \tag{34}$$

s' is the first derivative of s and δ is Euler–Poincaré characteristic of $\{s(x): x \in \mathscr{X}\}$ such that (1) $\delta = 1$ if $\mathscr{X} = [a, b]$ and $s(a) \neq s(b)$, and (2) $\delta = 0$ when $s(a) = s(b)$. Consult Kreyszig (1968) for more details on δ. Similarly, the coverage probability of two-sided SCR is

$$P(c) \approx \mathrm{pr}\left(\sup_{x\in\mathscr{X}} |W(x)| > c\right) \approx \frac{\kappa_0}{\pi}\exp(-c^2/2) + 2\delta\big(1-\Phi(c)\big).$$

Solution of C. By the symmetry of $W(x)$, it is easy to see that $c^l = c^u$. So, if c and $c^u = c^l$ are the solutions of

$$\alpha = \frac{\kappa_0}{\pi}\exp(-c^2/2) + 2\delta\big(1-\phi(c)\big), \quad \alpha = \frac{\kappa_0}{2\pi}\exp(-(c^l)^2/2) + \delta\big(1-\Phi(c^l)\big), \tag{35}$$

respectively, basic SCRs $l_c(x)$ in (2), $l^u(x)$ and $l^l(x)$ in (4) are $1-\alpha$ approximate two-sided, one-sided upper and lower, SCRs of $\eta(x)$ for all $x \in \mathscr{X}$.

REMARK 1. Note that a basic SCR for η or a SCR $g(l_c)$ for $E(y|x)$, based on c or c^u from the tube formula (35), is different from the naive SCR unless the responses in GLM are normally distributed. Using the naive SCR in discrete response cases could fail badly, as discussed in Section 4.

Our simulation results in the next section show that these basic SCRs based on the first level approximation in (35) are good if the sample size is large for many discrete models including logistic (for $n = 200$) and Poisson models (for $n = 50$). When n is small and the data are discrete, it is desirable to have a better approximation or correction which takes into account the bias from the Gaussian approximation in (18).

Corrections. There are two ways for modifying each of the two basic SCRs, based on (23) for one-sided basic SCR and based on (25) and (31) for two-sided basic SCR. We shall ignore the lattice term and name these corrections as follows:

Name	One-sided SCR	Two-sided SCR
Method 1 (based on a modified process)	Correction 1	Correction 2
Method 2 (bound a bias term)	Correction 1′	Correction 2′

Correction 1. Based on (23), define a new process

$$W_n^\sharp(x) = W_n(x) - \mu_n'(x) - \frac{\tau_n'(x)}{6}\{[W_n(x) - \mu_n'(x)]^2 - 1\}, \tag{36}$$

where $W_n(x) = (\hat{\eta}(x) - \eta(x))/\hat{\sigma}(x)$. It follows easily that $W_n^\sharp(x) \to^d W(x)$, by (18) and "$\mu_n' = O(n^{-1/2}) = \tau_n'$." Thus, we can define corrected upper and lower SCRs based on $W_n^\sharp$:

$$\begin{aligned} l^{u,1}(x) &= \{\eta(x)\colon W_n^\sharp(x) \ge -c^u\}, \\ l^{l,1}(x) &= \{\eta(x)\colon W_n^\sharp(x) \le c^l\} \qquad \forall\, x \in \mathscr{X}, \end{aligned} \tag{37}$$

where $c^u = c^l$ is a solution of (35).

Correction 1′. Our next corrected SCR is in the proposition below.

PROPOSITION 4.1. *Define* $R = \sup_{x\in\mathscr{X}} R(x)$ *where*

$$R(x) := \tfrac{1}{6}\tau_n'(x)[W(x)^2 - 1] + T_n(s(x)) + \mu_n'(x).$$

Suppose that there are positive constants $r = r_n$ *such that* $\mathrm{pr}(R \ge r) = o(\alpha)$ *as* $n \to \infty$ *and* $\alpha \to 0$. *Then if* $\mathscr{X}$ *is one-dimensional, a conservative and a liberal approximation to* (4) *are*

$$\begin{aligned} \alpha &\approx \frac{\kappa_0}{2\pi}\exp\left(-\frac{(c^u - r)^2}{2}\right) + (1 - \Phi(c^u - r))\delta, \\ \alpha &\approx \frac{\kappa_0}{2\pi}\exp\left(-\frac{(c^u + r)^2}{2}\right) + (1 - \Phi(c^u + r))\delta, \end{aligned} \tag{38}$$

respectively, where δ *is the same Euler–Poincaré characteristic as that in* (35). *Thus, our second corrected upper and lower one-sided SCRs are* l^u *and* l^l *in* (4) *with* $c^u = c^l$ *being the solutions from one of the equations in* (38).

The idea of the proof is to apply the inverse Edgeworth expansion (23) to approximate P^l in (24) and consider the $n^{-1/2}$ term, for example, R, as a bias term. Then the method of the bias correction from Sun and Loader (1994) is generalized to our problem for improving the accuracy of our approximation formula. See the Appendix for the proof.

Of course, the solution c^u depends on constants δ, κ_0 and r, which can be estimated. See (42)–(44) below. The resulting SCR l^u may have a slightly conservative [if the first equation in (38) is used] or a slightly liberal (if the second equation is used) coverage probability, but both coverage probabilities approach to $1-\alpha$ as $n \to \infty$ and then $\alpha \to 0$. See Section 5 for suggestions on choosing c^u from either of the two equations in (38) in practice.

Correction 2. Define new processes

$$\begin{aligned} W_n^{(1)}(x) &= |W_n(x)| + p_2(x, W_n(x)), \\ W_n^{(2)}(x) &= \left|W_n^{(0)}(x)\right| + q_2\left(x, W_n^{(0)}(x)\right). \end{aligned} \tag{39}$$

Then, similar to establishing (25) and (31), we have $W_n^{(1)}(x) =^d |W(x)| =^d W_n^{(2)}(x)$, up to $o_p(n^{-1})$. Thus, let c be the constant such that "$P\{\max_{\mathscr{X}} |W(x)| \le c\} = 1-\alpha$," that is, c be the solution of the first equation in (35), two corrected regions are

$$l^{(i)}(x) = \left\{\eta(x) \colon W_n^{(i)}(x) \le c\right\} \qquad \forall\, x \in \mathscr{X} \quad \text{for } i = 1, 2. \tag{40}$$

The SCR $l^{(1)}$ based on $W_n^{(1)}$ is no good, while $l^{(2)}$ based on $W_n^{(2)}$ provides a significant improvement over that based on W_n. See Batch 2 experiment in Section 5.

Correction $2'$. It is $l_c(x)$ in (2) with c being the solution of (41) below.

PROPOSITION 4.2. *Define* $R'_p = \sup_{x\in\mathscr{X}} p_2(x, W(x)) = O_p(1/n)$. *Suppose that there are positive constants* $r'_p = r'_p(n)$ *such that* $\mathrm{pr}(R'_p \le r'_p) = 1 - o(\alpha)$ *as* $n \to \infty$ *and* $n^{-k} \le \alpha \to 0$ *for some* $1 \ge k > 0$. *Then if* $\mathscr{X}$ *is one-dimensional, a conservative and a liberal approximation to* "$1 - \alpha = \mathrm{pr}\{\eta(x) \in l_c(x), \forall\, x \in \mathscr{X}\}$" *are*

$$\begin{aligned} \alpha &\approx \frac{\kappa_0}{\pi} \exp\left(-\frac{(c - r'_p)^2}{2}\right) + 2\left(1 - \Phi(c - r'_p)\right)\delta, \\ \alpha &\approx \frac{\kappa_0 - \kappa_1}{\pi} \exp\left(-\frac{(c + r'_p)^2}{2}\right) + 2\left(1 - \Phi(c + r'_p)\right)\delta, \end{aligned} \tag{41}$$

respectively, where δ *is same as* (35) *and* $\kappa_1 \approx 0.5 \int_{x\in\mathscr{X}_1} \|s'(x)\|\, dx$ *with* $\mathscr{X}_1$ *being the overlapping area of the tubular neighborhoods of* $\{s(x) \colon x \in \mathscr{X}\}$ *and* $\{-s(x) \colon x \in \mathscr{X}\}$.

REMARK 2. The corrected SCR 2′ from (41) are equivalent to

$$l^{(3)}(x) = \{\eta(x): |W_n(x)| \le c \pm r'_p\} \qquad \forall\, x \in \mathscr{X}.$$

The counterpart based on $W_n^{(0)}(x)$ is

$$l^{(4)}(x) = \{\eta(x): |W_n^{(0)}(x)| \le c \pm r'_q\} \qquad \forall\, x \in \mathscr{X},$$

where $\mathrm{pr}(R'_q \le r'_q) = 1 - o(\alpha)$ as $n \to \infty$ and $R'_q = \sup_{x\in\mathscr{X}} q_2(x, W(x)) = O_p(1/n)$. Interestingly, $l^{(3)}$ (based on $W_n^{(1)}$) works better than $l^{(4)}$ (based on $W_n^{(2)}$) in practice, reversing what occurred in Correction 2; see Batch 4 experiment in Section 5.

Estimation of constants $\kappa_0, \kappa_1, r, r'_p, r'_q$.

1. A simple approximation to κ_0 when $\mathscr{X} = [a, b]$ can be obtained by partitioning $[a, b]$ into $a = t_0 < \cdots < t_m = b$, and computing

$$(42)\quad \hat{\kappa}_0 = \sum_{i=1}^{m} \int_{t_{i-1}}^{t_i} \left\| \Delta \frac{(\hat{B}_n^T)^{-1} z(x)}{\sqrt{n}\hat{\sigma}(x)} \right\| dx \approx \sum_{i=1}^{m} \left\| \frac{(\hat{B}_n^T)^{-1} z(t_i)}{\sqrt{n}\hat{\sigma}(t_i)} - \frac{(\hat{B}_n^T)^{-1} z(t_{i-1})}{\sqrt{n}\hat{\sigma}(t_{i-1})} \right\|,$$

where $\hat{B}_n = B_n(\hat{\beta})$.

2. The calculation of κ_1 depends on the set $\mathscr{X}_1$, which can be approximated roughly by plotting $s(x)$ and $-s(x)$. As shown in Sun and Loader (1994), if $s: \mathscr{X} \to s(\mathscr{X})$ is 1–1, three times differentiable and there exists a vector γ with $\langle \gamma, s(x)\rangle > 0$ for all $x \in \mathscr{X}$, then the tubular neighborhoods of $\{s(x)\}$ and $\{-s(x)\}$ do not intersect for sufficiently small radii. The vector $\gamma = (1, \dots, 1)^T$ suffices in most regression models. In our simulations we simply set $\kappa_1 = 0$, which worked fine. A more rigorous and complicated expression for κ_1 may be derived using the inclusion and exclusion formulas due to Naiman and Wynn (1997). See the proof in the Appendix.

3. The estimate of r is more complicated than those of κ_0 and κ_1. The sharper it bounds R the better the correction will be. Analogous to (A.5) in the Appendix, a "sharp" estimate of r is

$$(43)\qquad \hat{r} = \hat{r}(c) = \sup_{x\in\mathscr{X}} \bigl(\tfrac{1}{6}\hat{\tau}'_n(x)[c^2 - 1] + \hat{T}_n(s_n(x)) + \hat{\mu}'_n(x)\bigr),$$

where c is a solution of an equation in (38) with r replaced by $\hat{r}(c)$. Thus, given a data set, a corrected conservative SCR 1′ is l_c with c to be the solution of the first equation in (38) where $r = \hat{r}(c)$. A "corrected" liberal SCR can be obtained similarly.

It is possible to have an estimate of r that does not depend on c, and hence it is easier than $\hat{r}$ in (43) to use for solving c based on (38) though it is not as sharp as $\hat{r}$. See Sun, Loader and McCormick (1998) for this estimate and a related interesting little inequality.

Using the same idea for developing $\hat{r}$, we have "sharp" estimates of r'_p and r'_q,

$$(44)\qquad \hat{r}'_p = \sup_x p_2(x, c) \quad \text{and} \quad \hat{r}'_q = \sup_x q_2(x, c).$$

5. Examples and simulations. To perform a simulation experiment we just need to compute the maxima (say, equal to b) of the processes $W_n(x)$, $W_n^{\sharp}(x)$, $W_n^{(0)}(x)$, $W_n^{(1)}(x)$ and $W_n^{(2)}(x)$ using the known true $\eta(x)$ and estimate the α quantile of a tube approximation $t(b)$ by

$$\hat{\alpha} = \sum_{i=1}^{10,000} \frac{\{t(b_i) < \alpha\}}{10,000}, \tag{45}$$

where b_i is the ith independent repetition of b in the experiment, 10,000 is the simulation size of the experiment and $t(\cdot)$ is a tube approximation to $P(\cdot)$ or $P^u(\cdot)$ or $P^l(\cdot)$ in (24). In other words, $t(b_i)$ is the right-hand side of an equation in (35) (for basic SCRs and Method 1, corrected SCRs 1 and 2), or the right-hand side of an equation in (38) or in (41) (for Method 2, corrected SCRs 1′ and 2′), evaluated at $c = b_i$ or $c^u = b_i$. If the SCR works well, $\hat{\alpha}$ should be close to the nominal noncoverage probability α. We will discuss our simulation results for one-sided and two-sided SCRs with $\alpha = 0.005$, 0.01, 0.05 and 0.1. The canonical (anti) links for the Poisson and Bernoulli–binomial regression are log and logistic links.

To actually compute SCRs for a data set, one places the critical values c or c^u obtained from the tube approximations on these processes to obtain equations that can be solved to find confidence limits for $\eta(x)$. See Table 3 below and an application to a real data set.

Simulations. Twenty x_i's were generated from $U[0, 1]$. For a given mean response function $E(Y_i|x_i) = m$, Poisson responses were generated, and a log-quadratic model fitted to the data. The basic two-sided SCRs $l_c(x)$ were computed over the interval [0, 1], and the coverage probabilities estimated based on 10,000 simulations by (45).

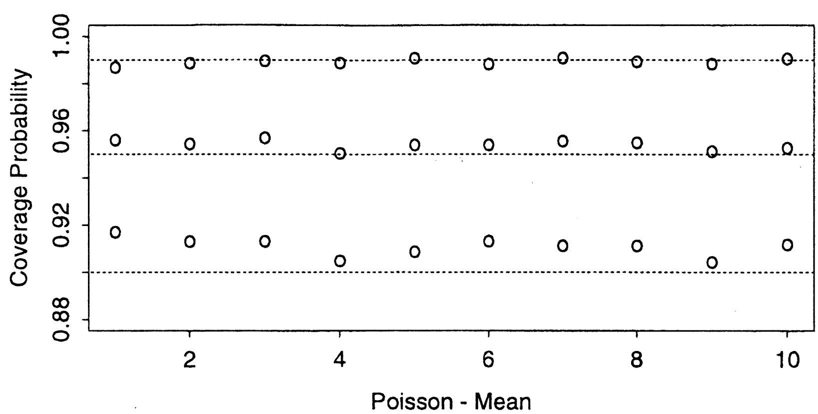

FIG. 1. *Simulated coverage probabilities for a Poisson log-quadratic regression. For each mean, the actual coverage obtained at a nominal* 90%, 95% *and* 99% *levels is displayed.*

Figure 1 displays the results. The coverage probabilities are generally quite good and comparable to what would be expected from the "naive" SCR for the mean response function from a Gaussian model. The "naive" SCR works reasonably well under a Gaussian model [Sun and Loader (1994)] but may fail under a nonnormal GLM (especially with discrete responses). Recall Remark 1 in Section 4. There is some evidence that the true distribution has light tails at smaller means, with the results being conservative at the nominal 90% coverage and slightly liberal at the nominal 99% coverage. However, the notable point is that the main effect (at the rate of $n^{-1/2}$) of skewness does not have a significant impact on coverage probability of two-sided SCRs, as shown in (16). Even though the distribution of $W_n(x)$ is quite heavily skewed [as shown below and in (22)], the main effects of skewness cancel at the two boundaries; the excess coverage gained at one boundary is lost at the other boundary.

To show the effect of skewness more dramatically, we consider one-sided confidence bounds. The coverage probabilities of one-sided basic SCRs l^u and l^l [based on $W_n(x)$] and corrected SCRs $l^{u,1}$ and $l^{l,1}$ [based on $W_n^\sharp(x)$] are shown in Figure 2. Clearly there is a dramatic skewness effect; The basic lower SCRs are liberal, and the basic upper SCRs are very conservative. The corrected SCRs improve the situation in both cases, although in the case of the upper bound, the correction does not go far enough (This is also true for the lower bound, since the tube formula should be slightly conservative, especially at the 90% level.) Part of the problem may be the use of estimated parameters in $\mu_n'(x)$ and $\tau_n'(x)$; however, the main problem appears to be a low kurtosis; $E((W_n(x) - \mu_n(x))^4 - (W_n(x) - \mu_n(x))^2) < 2$.

A second simulation is carried out in a Bernoulli setting, using quadratic logistic regression with true mean function 0.5 based on basic two-sided SCRs. Again, x_i's were generated from a $U[0, 1]$ distribution, with sample sizes ranging from 20 to 200. Results are shown in Figure 3. The results in this case are poor (in comparison to the results in the Poisson case), except for the largest sample size $n = 200$. Moreover, the results here cannot be attributed to bias

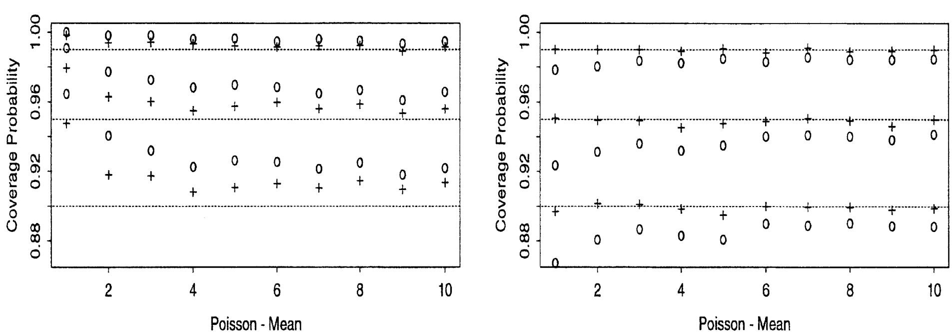

FIG. 2. *Simulated coverage probabilities for one-sided upper (left panel) and lower confidence bands (right panel). Coverages from the basic* SCRs *are shown with a* 0, *and coverages from a bias-skewness corrected* SCRs 1 *with a* +.

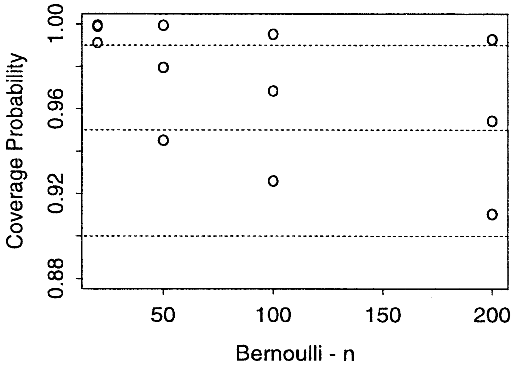

FIG. 3. *Simulated coverage probabilities for a quadratic logistic regression.*

and skewness; since our responses are Bernoulli with mean 0.5, symmetry implies the estimate is unbiased and has skewness 0. Therefore, Correction 1 and 1′ will not help with one-sided confidence bands in this case. The results with Corrections 2 and 2′ for two-sided SCRs are given below.

Implementing Corrections 2 and 2′ for two-sided SCRs is as straightforward as those for Corrections 1 and 1′, though the computation of the n^{-1} term is more complicated. Our design of experiments for comparing two-sided SCRs is a complete or partial crossing of the cases in Table 1 for the following four batches of experiments. In each experiment, simulation size is 10,000 and SCRs are computed for each of 10,000 independent samples from a distribution specified in Table 1, under an exact and/or overfitting scheme.

Batch 1. *Comparisons of overfits and exact fits*. The basic SCRs are computed from $W_n(x)$, for 10 pc models with deg $= 0, 1, 2$ fits and $n = 20, 50, 100$; for 2 bc models with deg $= 0, 1$ fits and for bl, bq, pl and pq with the exact fits and $n = 20, 50, 100, 150, 200$ and for Bl. This leads to a total of $90 + (4 +$

TABLE 1
Models and notations

Model	**Notation**	**Mean**	**Exact fit**
Poisson	pc	mean $= 1, 2, \dots, 10$	deg $= 0$
	pl	mean $= 5\exp(x)$, a log linear function	deg $= 1$
	pq	mean $= 10\exp(-1 + 6x - 9x^2)$, log quadratic	deg $= 2$
Bernoulli	bc	mean $= 0.5, 0.7$	deg $= 0$
	bl	mean $= \operatorname{expit}(x) := \exp(x)/(1 + \exp(x))$, logistic linear	deg $= 1$
	bq	mean $= \operatorname{expit}(x - 2x^2)$, logistic quadratic	deg $= 2$
Binomial	Bl	mean $= 3\ \operatorname{expit}(x)$	deg $= 1$

Sample size $n = 20, 50, 100, 150, 200$; overfits if fitted with deg $>$ exact deg.

4) $\times 5 + 1 = 131$ experiments. There are three points that can be made from these experiments.

The first point relates to the justification of overfits. In practice, we often do not know how a true mean response function depends on predictors. An underfit is much more serious than an overfit, since an underfitted model is wrong while an overfitted model still gives an unbiased estimator of the log (or logit) of the mean response function, though the variance of the estimator tends to be larger than that by an exact fit. An effect of this variance increase (the size of which decreases as the sample size increases) is that the SCR from an overfit tends to be more conservative. This is confirmed in this experiment; compare the noncoverage probabilities in Figure 4. Figure 4 only has the plot for $\alpha = 0.05$; the plots for other levels are similar. If there must be a deviation it is better to have a conservative SCR than a liberal SCR as long as the conservative SCR is informative—not too wide. In addition, Figure 4 shows interestingly that the coverage probabilities (not variances) under an overfitted model are *less varied* than those under an exact fit. Thus, Corrections $1'$ and $2'$ should work well in practice from an overfitted model. Of course, it is beneficial to perform an exact fit if possible (by performing diagnostics on several possible fits) before computing a SCR. This two-stage process, obtaining an exact fit to a given data and then computing a good SCR, will be illustrated in the data application below.

The second point compares the SCRs for Poisson and Bernoulli models, under exact and overfits. Most basic SCRs under the 90 pc cases are conservative, though some are a little liberal. Almost all SCRs under the remaining 41 cases are conservative. Generally, the basic SCRs from Bernoulli models are much more conservative than those from the Poisson models. This is consistent with our earlier discovery in Figures 1 and 3. Since the binomial model Bl is equivalent to a Bernoulli model bl with a larger sample size, this is the only batch where the Bl case is experimented.

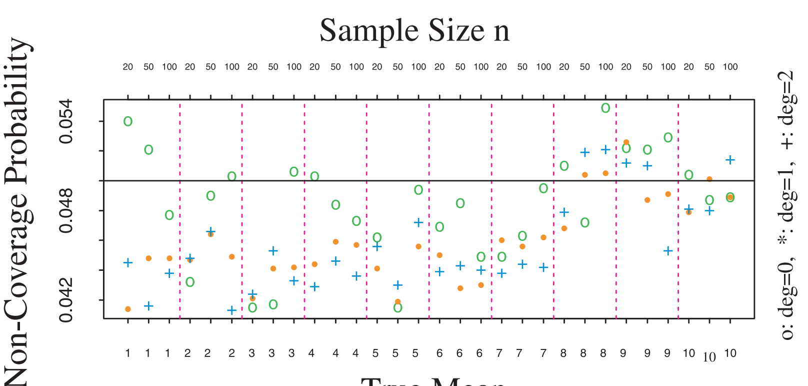

FIG. 4. *Comparisons of exact fits* (deg = 0) *and overfits* (deg = 1, 2). *Simulated noncoverage probabilities of the level* 0.05 *basic* SCRs *are plotted against* n *and the true mean in the Poisson model* pc.

The third point is about the computation of SCRs. In two out of these 131 experiments, our algorithm returned NA in one (or four) out of the 10,000 samples for the bc model with mean 0.7, $n = 20$ and a deg $= 0$ (or deg $= 1$) fit. This indicates that our algorithm occasionally has difficulties when the sample size is small, for example, $n = 20$.

The basic SCR is used as a baseline of comparisons in the rest of the experiments.

Batch 2. *Comparisons of $W_n^{(1)}(x)$ and $W_n^{(2)}(x)$ for building corrected SCR 2*. The SCRs computed from $W_n^{(2)}(x)$ are the clear winners, especially when the models are fitted exactly. This suggests performing exact fits and using a more centered process when possible.

Batch 3. *Comparisons of $W_n^{(0)}(x)$ and $W_n^{(2)}(x)$*. A SCR from $W_n(x)$ is the easiest to compute and that from $W_n^{(0)}(x)$ is easier to compute than that from $W_n^{(2)}(x)$; thus, if the results from two processes are not much different, we suggest using the simpler one. There were a total of 390 experiments with different fits, n and models. The results are shown in Figures 5 and 6 and summarized in Table 2, where by small α is meant $\alpha = 0.005, 0.01$.

For Bernoulli models bc, the improvement from $W_n^{(0)}(x)$ over $W_n(x)$ is not as much as $W_n^{(2)}(x)$, which is consistent with our earlier observation that the less satisfactory results of the basic SCRs "cannot be attributed to the main effects of skewness." The exceptions for the case of $n = 20$ when α is small is not surprising since a $1/n$ term is less predictable when the sample size is small. In fact, our `parfit` can fail to converge in some $n = 20$ cases. This is expected: although $\kappa_2(x) > 0$ in theory, κ_2 can be negative when it is estimated in a small sample case. Fortunately, the failure happened only when $n = 20$ and in the worst case only six out of 10,000 repetitions. We see similar results in the case of deg $= 1$ fits. For bq, the corrected SCRs for $\alpha = 0.1$ (large α) generally improve over the basic SCRs, but not enough sometimes. This is the case that Correction $2'$ helped.

Batch 4. *Comparisons of Corrections* 2 *and* $2'$. The sampling distributions are bl, bq, pl and pq. The fits are exact and $n = 20, 50, 100, 150, 200$. In each experiment, nine SCRs are computed while only four of them are listed in Table 2. Other unlisted SCRs include corrected SCR $2'$ based on $\max\{|W_n^{(0)}(x)|\} \pm \hat{r}'_q$ and those simplified from $W_n^{(1)}$ and $W_n^{(2)}$, for example, $\max\{|W_n^{(0)}(x)| + q_2(x, c)\}$ is simpler for building a SCR than $\max\{W_n^{(2)}(x)\} = \max\{|W_n^{(0)}(x)| + q_2(x, W_n^{(0)}(x))\}$. The unlisted SCRs though justified asymptotically [see (A.5)] did not perform well in simulations.

Again, the basic SCRs generally have reasonable coverage probabilities for Poisson data, but not for Bernoulli data unless the sample size is 200. The centered SCRs are uniformly better than the basic SCRs, though the improvements are not as big as corrected SCR 2 in some cases for $\alpha = 0.1$ and 0.05. Corrected SCRs 2 are excellent for pq models, very good for bl except $n = 20$ and pretty good for pl. Corrected SCRs $2'$ defined by $\{\eta(x)\colon |W_n(x)| \le c - |\hat{r}'_p|, x \in$

$\mathscr{X}\}$ (which serves as a guideline for deciding whether to add or subtract $\hat{r}_p$ in the confidence width) are better than corrected SCRs 2 for bl and bq.

EXAMPLE. The following is the insect data from Bliss (1935):

$\log_2$ (concentration)	0	1	2	3	4
No. of deaths	2	8	15	23	27
No. of insects	30	30	30	30	30

An exploratory data analysis suggests that the log2(concentration) of an insecticide might be a good predictor for modeling the probability of death of an insect. The analysis of a deviance table for a linear logistic fit and a quadratic logistic fit to the data [available in Sun, Loader and McCormick (1998)] suggests the linear logistic fit is good enough. This is also confirmed by other simple diagnostics plots. The estimates of the linear predictor $\eta(x)$

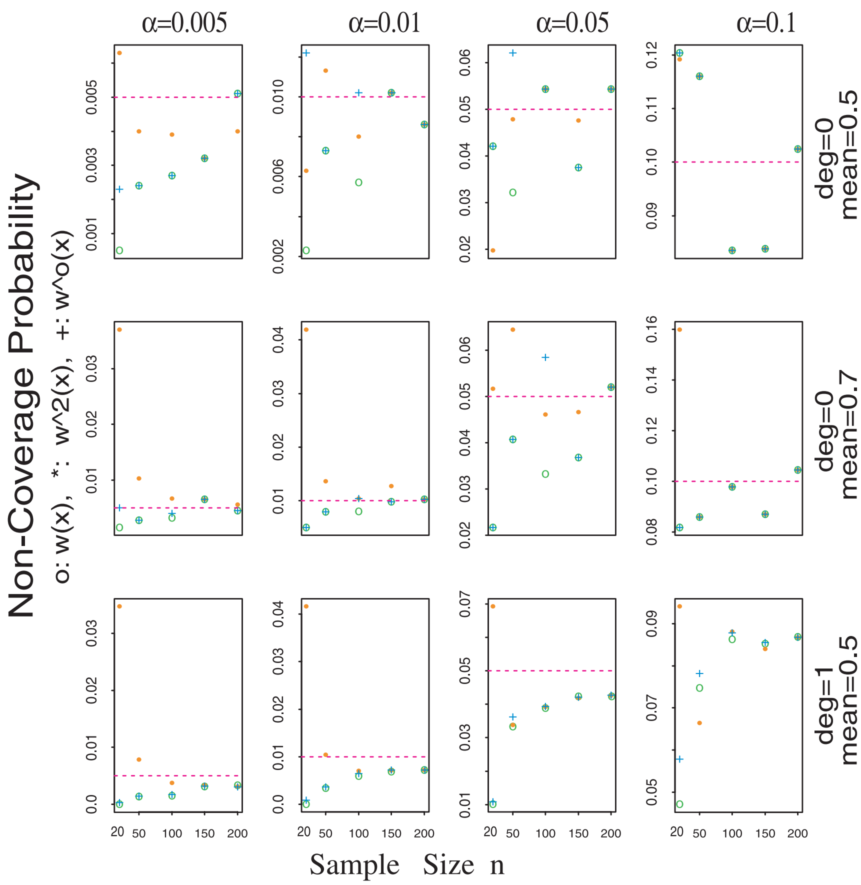

FIG. 5. *Comparisons of* $W_n(x)$, $W_n^{(2)}(x)$, $W_n^{(0)}(x)$, *for the Bernoulli models* bc.

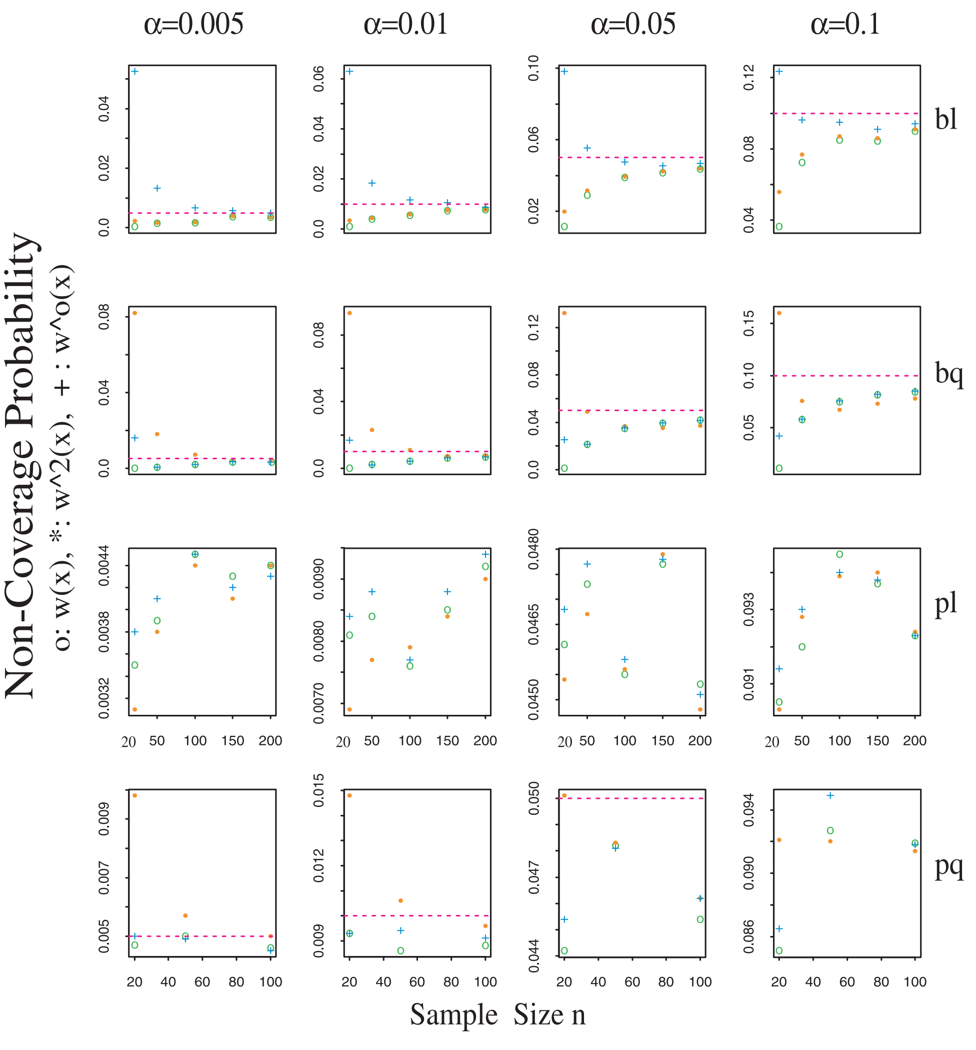

FIG. 6. *Comparisons of three SCRs with exact fits. For the* pq *model, results from three SCRs for* $n = 150$ *and* 200 *are very close and hence are omitted.*

TABLE 2
Comparisons of $W_n^{(0)}(x)$ and $W_n^{(2)}(x)$

	Winner	Exception
pc	$W_n^{(0)}(x)$	None
bc	$W_n^{(2)}(x)$	Use $W_n^{(0)}(x)$ if $n=20$
bl	$W_n^{(0)}(x)$	Use $W_n^{(2)}(x)$ if $n=20$ and $W_n(x)$ if $n=50$ when $\alpha=0.005, 0.01$
bq	$W_n^{(2)}(x)$	Use $W_n^{(0)}(x)$ if $n=20$ when $\alpha=0.005, 0.01$
pl	$W_n^{(0)}(x)$	Use $W_n(x)$ for $n>100$
pq	$W_n^{(0)}$ or W_n	Use $W_n^{(2)}(x)$ for $n=20$ when $\alpha=0.005, 0.01$

and its variance $\sigma^2(x)$ are

$$\hat{\eta}(x) = -2.32379 + 1.161895x,$$
$$\hat{\sigma}^2(x) = (1 \quad x)\begin{pmatrix} 0.17462 & -0.06582 \\ -0.06582 & 0.03291 \end{pmatrix}\begin{pmatrix} 1 \\ x \end{pmatrix},$$

where $x = \log_c(\text{concentration})$. Based on 41 equally spaced points in (0, 4) and $\alpha = 0.05$, we obtain $c = 2.4304$, $\hat{r}_p = 0.043414$. The summary statistics of $\hat{\kappa}_1(x)$, $\hat{\kappa}_2(x)$, $u(x)$ are

	Min	1st qu.	Median	**Mean**	3rd qu.	Max	st.dev.
$\hat{\kappa}_1(x)$	−0.036	−0.012	−0.00055	**−0.00011**	0.012	0.036	0.00279
$\hat{\kappa}_2(x)$	0.985	0.986	0.98660	**0.98670**	0.988	0.988	0.00015
$u(x)$	2.185	2.203	2.23700	**2.23800**	2.279	2.284	0.00517

Figure 7 displays an approximate pointwise confidence region (CR), $\hat{\eta}(x) \pm 1.96\hat{\sigma}(x)$, basic, centered, corrected simultaneous CRs in Table 3 for $E(Y|x) = p(x)$, the probability of death as a function of log2(concentration). The pointwise confidence region is narrowest of all CRs. This is expected since it can not

TABLE 3
Comparisons of corrected SCRs 2 and 2′ and recommendations

Name	**Process**	**SCR for $\eta(x)$**	**Recommendations**
Basic SCR	$\max\{\lvert W_n(x)\rvert\}$	$\hat{\eta}(x) \pm c\hat{\sigma}(x)$,	For* pc, pq
Centered SCR	$\max\{\lvert W_n^{(0)}(x)\rvert\}$	$\hat{\eta}(x) - \hat{\kappa}_1(x)\hat{\sigma}(x) \pm c\hat{\sigma}(x)\sqrt{\hat{\kappa}_2(x)}$	For* pc, bl, pl
Corrected SCR 2	$\max\{W_n^{(2)}(x)\}$	$\hat{\eta}(x) \pm u(x)\hat{\sigma}(x)$	For* bc, bq, pl, pq
Corrected SCR 2′	$\max\{\lvert W_n(x)\rvert\} \pm \hat{r}'_p$	$\hat{\eta}(x) \pm (c - \lvert\hat{r}_p\rvert)\hat{\sigma}(x)$	For bl, bq

Here c is the first level approximation from the first equation in (35), $\hat{r}'_p$ is in (44), $\hat{\kappa}_1(x)$ and $\hat{\kappa}_2(x)$ are estimated mean and variance of $W_n(x)$ in (28), and $u(x)$ is the solution of "$|u| + q_2(x, u) = c$," with q_2 defined in (32).

*Summarized based on Table 2 and Batch 4 experiments: briefly, if bl and pl, use corrected SCR 2′; if one is to choose between the first three SCRs, follow Table 2 for exceptions and specifics related to n and α.

be interpreted simultaneously for making inferences about the entire curve as the other SCRs do. All SCRs are fairly close since the sample size (number of insects) is 150. However, the basic SCR is the widest (as expected) and centered SCR is indistinguishable from the basic SCR. Both corrected SCR 2 and 2′ are narrower than the basic SCR with the corrected SCR 2 being the best.

6. Discussion and concluding remarks. We constructed basic SCRs and corrected SCRs for the mean response function in a GLM. The central idea in our approach is similar to the Skorohod construction so that we could explicitly study the "bias" from approximating $W_n(x)$ by $W(x)$. When this bias is small we can estimate the bias for adjusting the approximating formula.

Basic SCRs are informative [an original Figure 8 is omitted. If interested, see Sun, Loader and McCormick (1998)]. They are reasonable for continuous data and Poisson data when $n \geq 50$ and binomial data when $n \geq 200$. In other cases, centered or corrected SCRs should be used. Our centered SCRs uniformly improve over the basic SCRs and corrected SCRs in most cases offer improvement over the basic SCRs. Corrections 1 are generally helpful for one-sided bands. Corrections 2 and 2′ can be used for two-sided bands. See Table 3 for a summary of recommended two-sided SCRs. Implementing these SCRs via our software `parfit` is fairly straightforward.

Our methods require that the link function be known. When the link function is unknown, a user may use residual plots to guess a reasonable link function and then proceed as above as if the link were known. In this case,

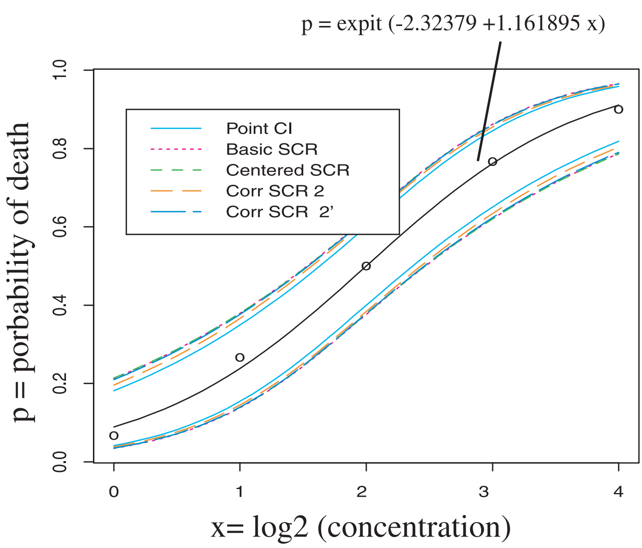

FIG. 7. *Approximate pointwise* CI*, basic, centered, corrected simultaneous* CI *for* $E(Y|x) = p(x)$. *The points are percentages of deaths at each* $\log 2$ (*concentration*). *The middle black solid curve is the estimate* $\hat{p}(x) = \text{expit}(-2.32379 + 1.161895 * x)$.

a further justification may be needed to justify the use of the plug-in link function.

Alternative methods for constructing a SCR can be found in Knafl, Sacks and Ylvisaker (1985), Hall and Titterington (1988), Härdle and Marron (1991) and others. Knafl, Sacks and Ylvisaker (1985) assume that the errors are normal and additive. They build simultaneous confidence bands on discrete points using discrete upcrossings and then interpolate between these points, so in the limit it is equivalent to the "tube method" in the one-dimensional case ($d = 1$) when errors are normal. Härdle and Marron (1991)'s bands are based on kernel estimates of the regression function and a bootstrap procedure for building simultaneous error bars, that is, their bands are simultaneous on a set of finite discrete points in the predictor space. Some comparisons with the tube formula approach are presented in Loader (1993). Hall and Titterington (1988) present an alternative approach for constructing SCRs for density functions and regression functions. They divide the data into nonoverlapping bins and then use the data in each bin to find confidence intervals at the centers of each bin (this essentially amounts to using a rectangular kernel estimate). Since the bins are nonoverlapping, the intervals for separate bins are independent, so they can easily make these simultaneous. Then, to construct a SCR, for a regression problem, say, they assume a bound on a low-order derivative of the regression function. So this bound plays an important role in the actual size and coverage probability of their SCR. In any case, the tube formula approach is complementary to previous bands [cf. Sun and Loader (1994) and Loader (1993)].

APPENDIX
DERIVATION AND EXAMPLES

Exemplar μ_n, τ_n and T_n's.

EXAMPLE 1 (Gamma). Here the response variables have a gamma density with a known shape parameter $r > 0$ and an unknown scale parameter $\omega > 0$,

$$f(y; r, \omega) = \frac{\omega^r}{\Gamma(r)} y^{r-1} e^{-\omega y} \quad \text{for } y > 0.$$

We have the form of (1) with $\theta = -\omega$, $b(\theta) = -r\log(-\theta)$. Clearly, $g(\theta) = b'(\theta) = -r/\theta$ is monotone in θ, $-\omega_i = \theta_i = \langle z(x_i), \beta\rangle$ and

$$b'(\theta) = -\frac{r}{\theta} = \frac{r}{\omega}, \quad b''(\theta) = \frac{r}{\theta^2} = \frac{r}{\omega^2}, \quad b^{(3)}(\theta) = -\frac{2r}{\theta^3} = \frac{2r}{\omega^3}.$$

Substituting them into (13) and (14) etc. gives

$$A_n = \frac{1}{n}\sum_{i=1}^{n} \frac{r}{\theta_i^2} z_i z_i^T, \quad \mu_n = \frac{1}{n^{3/2}}\sum_{i=1}^{n} \frac{r}{\theta_i^3}\|u_i\|^2 u_i \quad \text{and} \quad \tau_n(v) = \frac{4}{n^{3/2}}\sum_{l=1}^{n} \frac{r}{\theta_i^3}\langle u_i, v\rangle^3.$$

In this case ξ *cannot* be lattice as Y_i is a continuous random variable and hence T_n in Proposition 3.1 is zero.

Note that a commonly used *noncanonical* link for the gamma case is log-link, that is,

$$\log(\omega) = \langle z(x), \beta\rangle.$$

In this case, a similar Edgeworth expansion exists though the maximum likelihood estimates of β, τ_n and μ_n may be different from ones under the canonical link.

EXAMPLE 2 (Poisson). Here the response variables have a Poisson density with an intensity parameter $\omega > 0$,

$$f(y;\omega) = \frac{\omega^y}{y!}e^{-\omega} \quad \text{for } y = 0, 1, \dots.$$

We have the form of (1) with $\theta = \log(\omega)$, $b(\theta) = e^\theta$. Again, $g(\theta) = b'(\theta) = e^\theta$ is monotone in θ, and $\log(\omega_i) = \theta_i$, so $b'(\theta) = b''(\theta) = b^{(3)}(\theta) = e^\theta = \omega$. Substituting them into (13) and (14) etc., yields similar expressions of A_n, μ and τ_n.

Note that ξ_j has the lattice distribution iff the jth component of $\sum_{i=1}^n (y_i - \omega_i) z_i$ does, or if the jth component of z_i for all i only take lattice values; that is, they belong to $a_j + h_j\mathscr{Z}$ for some a_j, h_j, for $j = 1, \dots, q$, where $\mathscr{Z}$ is the integer set. *Hence,* $T_n = 0$ *unless we have a very special design, for example,* $z_i^T = (1, \dots, 1)$, *which is rare.*

EXAMPLE 3 (Binomial). Here the response variables have a binomial density with a success parameter $0 < p < 1$,

$$f(y;p) = \binom{n}{y} p^y (1-p)^{n-y} \quad \text{for } y = 0, 1, \dots, n.$$

The GLM with such a random structure is called a logistic model. We also have the form of (1) with $\theta = \log(p/(1-p))$, $b(\theta) = n\log(1+e^\theta)$. So $g(\theta) = b'(\theta)$, and for $q = 1 - p$,

$$b'(\theta) = \frac{ne^\theta}{1+e^\theta} = np, \qquad b''(\theta) = \frac{ne^\theta}{(1+e^\theta)^2} = npq,$$

$$b^{(3)}(\theta) = \frac{ne^\theta}{(1+e^\theta)^2} - \frac{n2e^{2\theta}}{(1+e^\theta)^3} = npq - 2np^2q.$$

Note that ξ might have a lattice distribution if we use the same special design described in Example 2: $z_i^T = (1, \dots, 1)$.

In most cases, $\langle v, \xi\rangle$ *does not have a lattice* distribution, and it is approximately normally distributed.

PROOF OF PROPOSITION 3.1 AND EQUATION (11). Let β_0 be the true value of β, set $\theta_{i,0} = z(x_i)^T\beta_0$, and denote

$$\psi_n := \frac{1}{n}\sum_{i=1}^n (Y_i - b'(\theta_{i,0}))z_i \quad \text{and} \quad g_n(\beta) := \frac{1}{n}\sum_{i=1}^n \big(b'(\theta_i) - b'(\theta_{i,0})\big)z_i.$$

Then from (6) we see that $\hat{\beta}$ is a solution to

$$\psi_n = g_n(\beta). \tag{A.1}$$

Note that (A.1) has as a first-order approximation the equation

$$\psi_n = \frac{1}{n}\sum_{i=1}^{n} b^{(2)}(\theta_{i,0})\langle z_i, \beta - \beta_0\rangle z_i = A_n(\beta - \beta_0)$$

with solution $\hat{\beta}_{n,1}$ given by "$\hat{\beta}_{n,l} - \beta_0 = A_n^{-1}\psi_n$." A second-order approximation to (A.1) yields the equation

$$\psi_n = \frac{1}{n}\sum_{i=1}^{n} b^{(2)}(\theta_{i,0})\langle z_i, \beta - \beta_0\rangle z_i + \frac{1}{2n}\sum_{i=1}^{n} b^{(3)}(\theta_{i,0})\langle z_i, \beta - \beta_0\rangle^2 z_i.$$

Replacing $\beta - \beta_0$ with the first-order solution $A_n^{-1}\psi_n$ in the second term on the right-hand side above yields a linear equation with solution $\hat{\beta}_{n,2}$. Continuing in this way one can obtain an expansion for $\hat{\beta}_n$ defined in (A.1) in terms of ψ_n. Taking the branch that sets $\hat{\beta}_n = \beta_0$ when $\psi_n = 0$ yields

$$\hat{\beta}_n - \beta_0 = A_n^{-1}\psi_n - \frac{1}{2n}\sum_{i=1}^{n} b^{(3)}(\theta_{i,0})\Big[z(x_i)^T A_n^{-1}\psi_n\Big]^2 A_n^{-1} z(x_i) + o_p(|\psi_n|^2),$$

that is, $\hat{\beta}_n = \hat{\beta}_{n,2}$ if $o_p(|\psi_n|^2)$ term above is zero. This is equivalent to

$$B_n\sqrt{n}(\hat{\beta}_n - \beta_0) = \xi - \frac{1}{2n^{3/2}}\sum_{i=1}^{n} b^{(3)}(\theta_{i,0})\langle u_i, \xi\rangle^2 u_i + o_p(n^{1/2}|\psi_n|^2). \tag{A.2}$$

Note that $\psi_n = n^{-1/2}B_n^T\xi = O_p(n^{-1/2})$ with ξ defined in (12), so (A.2) implies (11).

The moments of ξ can be calculated easily as

$$E\langle a_0, \xi\rangle = 0, \qquad E\langle a_0, \xi\rangle\langle a_1, \xi\rangle = \langle a_0, a_1\rangle,$$

$$E\langle a_0, \xi\rangle\langle a_1, \xi\rangle\langle a_2, \xi\rangle = \frac{1}{n^{3/2}}\sum_{l=1}^{n} b^{(3)}(\theta_l)\langle a_0, u_l\rangle\langle a_1, u_l\rangle\langle a_2, u_l\rangle,$$

$$\begin{aligned} E\langle a_0, \xi\rangle\langle a_1, \xi\rangle\langle a_2, \xi\rangle\langle a_3, \xi\rangle = {} & \langle a_0, a_1\rangle\langle a_2, a_3\rangle \\ & + \langle a_0, a_2\rangle\langle a_1, a_3\rangle + \langle a_0, a_3\rangle\langle a_1, a_2\rangle, \\ & + \frac{1}{n^2}\sum_{l=1}^{n}\big[b^{(4)}(\theta_l) + 3b''(\theta_l)\big] \\ & \qquad \times \langle a_0, u_l\rangle\langle a_1, u_l\rangle\langle a_2, u_l\rangle\langle a_3, u_l\rangle \end{aligned} \tag{A.3}$$

for any nonrandom vectors $a_0, \ldots, a_3$. It then follows from (A.2) and (A.3) that the bias and skewness of $\langle v, B_n\sqrt{n}(\hat{\beta}_n - \beta)\rangle$ are $\langle v, \mu_n\rangle$ and $\tau_n(v)$ in (13) and (14), respectively.

In the nonlattice case it follows again from (A.2) that for Φ and ϕ denoting the standard normal distribution and density, respectively,

$$\begin{aligned} &\Big|\mathrm{pr}\big\{\langle v, B_n\sqrt{n}(\hat{\beta}_n - \beta_0) - \mu_n\rangle \le x\big\} \\ &\qquad -\Phi(x) - \frac{1}{6}\tau_n(v)(1 - x^2)\phi(x)\Big| = o\Big(\frac{1}{\sqrt{n}}\Big). \end{aligned} \tag{A.4}$$

Taking note of

$$\left|\mathrm{pr}\left\{Z+\frac{1}{6}\tau_n(v)[Z^2-1]\le x\right\}-\mathrm{pr}\left\{Z+\frac{1}{6}\tau_n(v)[x^2-1]\le x\right\}\right|=o\left(\frac{1}{\sqrt{n}}\right),\tag{A.5}$$

where $\tau_n(v)=O(1/\sqrt{n})$ and

$$\left|\mathrm{pr}\left\{Z\le x+\frac{1}{6}\tau_n(v)(1-x^2)\right\}-\Phi(x)-\frac{1}{6}\tau_n(v)(1-x^2)\phi(x)\right|=o\left(\frac{1}{\sqrt{n}}\right),$$

we obtain

$$\left|\mathrm{pr}\left\{Z+\frac{1}{6}\tau_n(v)[Z^2-1]\le x\right\}-\Phi(x)-\frac{1}{6}\tau_n(v)[1-x^2]\phi(x)\right|=o\left(\frac{1}{\sqrt{n}}\right);\tag{A.6}$$

thus we see from (A.4) and (A.6) that up to order $o_p(1/\sqrt{n})$,

$$\langle v, B_n\sqrt{n}(\hat{\beta}_n-\beta_0)-\mu_n\rangle\stackrel{d}{=}Z+\frac{1}{6}\tau_n(v)[Z^2-1].$$

The inverse Edgeworth expansion now follows. The lattice case proceeds similarly, but using the notations in Hall [(1992), page 46]. □

PROOF OF (21) AND (22). For $\hat{A}_n=A_n(\hat{\beta})$, since

$$\begin{aligned}B_n\hat{A}_n^{-1}B_n^T&=B_n(A_n+\hat{A}_n-A_n)^{-1}B_n^T=\left[I+(B_n^T)^{-1}(\hat{A}_n-A_n)B_n^{-1}\right]^{-1}\\&=I-(B_n^T)^{-1}(\hat{A}_n-A_n)B_n^{-1}+o_p(\hat{A}_n-A_n)\\&=I-\frac{1}{n}\sum_{i=1}^n u_iu_i^T(b''(\hat{\theta}_i)-b''(\theta_i))+o_p(\hat{A}_n-A_n)\quad\text{[by (8)]}\\&=I-\frac{1}{n^{3/2}}\sum_{i=1}^n b^{(3)}(\theta_i)u_iu_i^T\langle u_i,\xi\rangle(1+o_p(1))\quad\text{[by (A.2)]}\end{aligned}$$

we have for $v=(B_n^T)^{-1}z(x)$ that $s_n(x)=v/\|v\|$ and (omitting the subscripts of s)

$$\begin{aligned}\frac{\sigma(x)}{\hat{\sigma}(x)}&=\left[\frac{z(x)^TA_n^{-1}z(x)}{z(x)^T\hat{A}_n^{-1}z(x)}\right]^{1/2}=\frac{\|v\|}{\left[v^TB_n\hat{A}_n^{-1}B_n^Tv\right]^{1/2}}\\&=\left[1-\frac{1}{n^{3/2}}\sum_{i=1}^n b^{(3)}(\theta_i)\left\langle\frac{v}{\|v\|},u_i\right\rangle^2\langle u_i,\xi\rangle(1+o_p(1))\right]^{-1/2}\\&=1+\frac{1}{2n^{3/2}}\sum_{i=1}^n b^{(3)}(\theta_i)\langle s(x),u_i\rangle^2\langle u_i,\xi\rangle+o_p(n^{-1/2}).\end{aligned}$$

Hence, from (A.2) again,

$$
\begin{aligned}
W_n(x) = \frac{\langle s(x), B_n\sqrt{n}(\hat{\beta} - \beta)\rangle\sigma(x)}{\hat{\sigma}(x)} &= \langle s(x), \xi\rangle \\
+\frac{1}{2n^{3/2}}\sum_{i=1}^{n} b^{(3)}(\theta_i)\big[&\langle s(x), u_i\rangle^2\langle u_i, \xi\rangle\langle s(x), \xi\rangle \\
&- \langle u_i, s(x)\rangle\langle u_i, \xi\rangle^2\big] + o_p(n^{-1/2}),
\end{aligned}
$$

which is (21). Equations in (22) then easily follow from (A.3). □

PROOF OF PROPOSITION 4.1. By the inverse Edgeworth expansion (23),

$$
P^l(c) \approx \mathrm{pr}\Big(\sup_{x\in\mathscr{X}}[W(x) + R(x)] > c\Big) \le \mathrm{pr}\Big(\sup_{x\in\mathscr{X}} W(x) + R > c\Big),
$$

where $R = O_p(n^{-1/2})$ and P^l is in (24). Let r_n be positive constants for which $P\{R_n \ge r_n\} = o(\alpha)$ as $n \to \infty$. Then, similar to establishing the first level approximation (35) to the basic SCR, we have

$$
\begin{aligned}
P^l(c) &\le \mathrm{pr}\Big\{\sup_{x\in\mathscr{X}} W(x) > c - r\Big\} + o(\alpha) \\
&\approx \frac{\kappa_0}{2\pi}\exp\Big(-\frac{(c-r)^2}{2}\Big) + (1 - \Phi(c-r)) + o(\alpha).
\end{aligned}
$$

Setting the right-hand side equal to α gives the first equation of (38). This is a conservative confidence band, but the nominal level will be closely approximated if r is close to R and α is small. An approximate lower bound to $P^l(c)$ can be obtained similarly, which leads to (38). □

PROOF OF PROPOSITION 4.2. This is similar to that of Proposition 4.1. The main differences are as follows. For the upper bound approximation, we note that

$$
\begin{aligned}
P(c) &= \mathrm{pr}\Big(\sup_{x\in\mathscr{X}}\{|W(x)| - p_2(x, W(x))\} > c\Big) + o(\alpha) \\
&\le \mathrm{pr}\Big(\sup_{x\in\mathscr{X}} |W(x)| + R'_p > c\Big) + o(\alpha) \\
&\le 2\,\mathrm{pr}\Big\{\sup_{x\in\mathscr{X}} W(x) > c - r'_p\Big\} + o(\alpha).
\end{aligned}
$$

Setting the last formula equal to α gives the first equation of (41). For the lower bound approximation, note that for $c' = c - r'_p$,

$$\begin{aligned} P(c) &\geq \mathrm{pr}\left(\sup_{x \in \mathscr{X}} |W(x)| - R'_p > c\right) + o(\alpha) \\ &\geq \mathrm{pr}\left\{\sup_{x \in \mathscr{X}} |W(x)| > c'\right\} + o(\alpha) \\ &= \int_{c'}^{\infty} \mathrm{pr}\left\{\sup_{x \in \mathscr{X}} |\langle s(x), U\rangle| \geq \frac{c'}{y}\right\} g(y, n)\, dy + o(\alpha), \end{aligned}$$

where $g(y, n)$ is the density of a χ random variable with n degrees of freedom and U is a uniform random variable on the unit sphere S^{n-1}.

Let $\mathscr{T}(r|\mathscr{M})$ be a tube of $\mathscr{M} = \{s(x), x \in \mathscr{X}\}$ with radius r (by abusing the notation r used earlier to indicate another constant: radius),

$$\mathscr{T}(r|\mathscr{M}) = \{u\colon u \in S^{n-1}, \inf_{x \in \mathscr{X}} \|u - s(x)\| \leq r\}.$$

Then, in the one-sided case,

$$P\left\{\sup_{x \in \mathscr{X}} \langle s(x), U\rangle \geq \frac{c'}{y}\right\} = P\{U \in \mathscr{T}(r|\mathscr{M})\}$$

by the simple equality $\|a - b\|^2 = 2 - 2\langle a, b\rangle$ where $a, b \in S^{n-1}$ and $r = \sqrt{2(1 - c'/y)}$. In the two-sided case,

$$P\left\{\sup_{x \in \mathscr{X}} |\langle s(x), U\rangle| \geq \frac{c'}{y}\right\} = P\{U \in \{\mathscr{T}(r|\mathscr{M}) \cup \mathscr{T}(r|-\mathscr{M})\}\}.$$

The probability on the right-hand side is equal to the ratio of the volume of the tubular neighborhood $\mathscr{T}(r|\mathscr{M}) \cup \mathscr{T}(r|-\mathscr{M})$ and the volume of the unit sphere S^{n-1}. By applying Weyl's (1939) formula, as in Sun (1993), for the volume of the tubular neighborhood and ignoring boundary effects, we have

$$\begin{aligned} \int_{c'}^{\infty} \mathrm{pr}\left\{\sup_{x \in \mathscr{X}} |\langle s(x), U\rangle| \geq \frac{c'}{y}\right\} g(y, n)\, dy &\approx \int_{c'}^{\infty} (2\kappa_0 - \kappa_1(y)) J_0\left(\frac{c'}{y}\right) g(y, n)\, dy \\ &\approx \frac{2\kappa_0 - \kappa_1}{2\pi} \exp\left(-\frac{c'^2}{2}\right), \end{aligned}$$

where κ_0 is in (34) and

$$\begin{aligned} J_0(w) &= \frac{\Gamma(n/2)}{\Gamma((n-2)/2)\pi} \int_w^1 (1 - t^2)^{(n-4)/2} t\, dt, \\ \kappa_1 &= \pi \exp\left(\frac{c'^2}{2}\right) \int_{c'}^{\infty} \kappa_1(y) J_0\left(\frac{c'}{y}\right) g(y, n)\, dy, \\ \kappa_1(y) &= \int_{\mathscr{X}_1(y)} \|s'(x)\|\, dx, \\ \mathscr{X}_1(y) &= [x \in \mathscr{X}\colon \mathscr{T}(r|\mathscr{M}) \cap \mathscr{T}(r|-\mathscr{M})], \qquad r = 2(1 - c'/y). \end{aligned}$$

If $\kappa_1(y)$ is independent of y for $y \le C$, some constant, then

$$\kappa_1 = \int_{\mathscr{X}_1} \|s'(x)\|\, dx, \qquad \mathscr{X}_1 = \{s(x),\ x \in \mathscr{X}\} \cap \{-s(x),\ x \in \mathscr{X}\}.$$

This is true for most regression problems. Also recall the discussion on computing κ_1 in Section 4. In summary, the addition of κ_1 is to make sure that the intersection of $\mathscr{M}$ and $-\mathscr{M}$ is only counted once in computing the volume. Finally, adding the boundary correction $2\delta(1-\Phi(c))$ as that in Sun and Loader (1994) leads to the liberal formula (41). □

Notation.

n	sample size
x_i	ith predictor
$\mathscr{X}$	domain of interest about x_i
y_i or Y_i	ith response variable
g	link function, $g(\eta(x_i)) = E(Y_i \mid x_i)$
f	density of Y_i; see (1)
$a(y)$	component of f; see (1)
$b(\theta)$	component of f; see (1)
θ	natural parameter; see (1)
$\theta_i, z_i \dots$	subscripted functions, denoting functions evaluated at x_i, e.g., $z_i = z(x_i)$, $\theta_i = \theta(x_i)$
$\eta(x)$	linear predictor, $\eta(x) = z(x)^T\beta$
β	unknown q-dimensional parameter
$z(x)$	known q-dimensional covariate
$\hat{}$	denote estimate or estimator, e.g., $\hat\beta$ is an estimator of β
l or $l_c(x)$	two-sided basic SCR in (2)
$l^u(x)$	one-sided basic upper SCR in (4)
$l^l(x)$	one-sided basic lower SCR in (4)
$\hat\sigma^2(x)$	estimated variance of $\hat\eta$; see (7)
$I_n(\beta)$	$q \times q$ Fisher information matrix; see (8)
X	$n \times q$ design matrix: $X^T = (z_1, \dots, z_n)$; see (8)
Σ	$n \times n$ covariance matrix; see (8)
B_n	upper Cholesky triangle, $B_n^T B_n = A_n$
A_n	$A_n = A_n(\beta) = I_n(\beta)/n$
w_i	$w_i = A_n^{-1} z_i$
u_i	$u_i = (B_n^T)^{-1} z_i$
S^{d-1}	unit sphere, $S^{d-1} = \{v \in \mathbb{R}^d : \|v\| = 1\}$
$\|\cdot\|$	Euclidean norm, $\|v\|^2 = \sum_{i=1}^d v_i^2$
$\xrightarrow{d}$	convergence in distribution; see the second paragraph in Section 3

$\stackrel{d}{=}$	equality of distributions; see the second paragraph in Section 3
ξ	normalized vector; see (12)
μ or μ_n	bias; see (13)
τ_n	skewness; see (14)
$T_n(\nu)$	lattice term; see (15)
λ_n	eigenvalues of A_n; see Proposition 3.1
h or H	periodic function; see (15)
$W_n(x)$	$W_n(x) = (\hat{\eta}(x) - \eta(x))/\hat{\sigma}(x)$; see (17)
$P(c)$	$P(c) = \mathrm{pr}\{\sup_{x\in\mathscr{X}} \lvert W_n(x)\rvert > c\}$; see (24)
$P^l(c)$	$P^l(c) = \mathrm{pr}\{\sup_{x\in\mathscr{X}} W_n(x) > c\}$; see (24)
$P^u(c)$	$P^u(c) = \mathrm{pr}\{\inf_{x\in\mathscr{X}} W_n(x) < -c\}$; see (24)
$W(x)$	limit process of $W_n(x)$; see (18)
$s(x)$	$s(x) = \lim_{n\to\infty} s_n(x)$
$s_n(x)$	$s_n(x) = (B_n^T)^{-1}z(x)/(\sqrt{n}\sigma(x))$; see (19)
$\Phi(\cdot)$	standard normal cumulative distribution
κ_0	volume of $\{s(x)\colon x \in \mathscr{X}\}$; see (34) and (42)
$W_n^*(x)$	$W_n^*(x) = (\hat{\eta}(x) - \eta(x))/\sigma(x)$; see (20) and compare it with $W_n(x)$
μ_n'	bias of W_n; see (22)
τ_n'	skewness of W_n; see (22)
$W_n^\sharp(x)$	see (36)
$l^{u,1}(x)$	one-sided corrected upper SCR; see (37)
$l^{l,1}(x)$	one-sided corrected lower SCR; see (37)
R or $R(x)$	see Proposition 4.1
r, $\hat{r}$	see Proposition 4.1 and (43)
κ_1	see the paragraph before (43)
$p_2(x, z)$	see (27)
$\kappa_i(x)$	$i = 1, \dots, 4$; see (28)
$C_1 \dots C_9$	see (29)
$W_n^{(1)}(x)$	$W_n^{(1)}(x) = \lvert W_n(x)\rvert + p_2(x, W_n(x))$; see (39)
$W_n^{(2)}(x)$	$W_n^{(2)}(x) = \lvert W_n^{(0)}(x)\rvert + q_2(x, W_n^{(0)}(x))$; see (39)
$W_n^{(0)}(x)$	$W_n^{(0)}(x) = (W_n(x) - \hat{\kappa}_1(x))/\sqrt{\hat{\kappa}_2(x)}$; see (30)
$q_2(x, z)$	see (32)
$\kappa_{2,2}^s$, $\kappa_{3,1}^s$	see (33)
$\kappa_{4,1}^s$	see (33)
$l^{(1)}(x)$	see (40)
$l^{(2)}(x)$	see (40)
$l^{(3)}(x)$	see Remark 2
R', r', $\hat{r}'$	see Proposition 4.2 and (44)
r'', $\hat{r}''$	see Remark 2

Acknowledgments. The authors thank two referees and an associate editor for making this paper much more reader friendly and a referee for pointing out that the correction term for two-sided bands should be of order n^{-1} rather than $n^{-1/2}$. Jiayang Sun acknowledges the helpful discussion she had with Bob Keener and Dan Rabinowitz at the initial stage of this research.

REFERENCES

BLISS, C. L. (1935). The calculation of the dosage-mortality curve. *Ann. Appl. Biology* **22** 134–167.

BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families: with Applications in Statistical Decision Theory*. IMS, Hayward, CA.

FARAWAY, J. and SUN, J (1995). Simultaneous confidence bands for linear regression with heteroscedastic errors. *J. Amer. Statist. Assoc.* **90** 1094–1098.

HALL, P. and TITTERINGTON, D. M. (1988). On confidence bands in nonparametric density estimation and regression. *J. Multivariate Anal.* **27** 228–254.

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer, New York.

HÄRDLE, W. and MARRON, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* **19** 778–796.

JOHANSEN, S. and JOHNSTONE, I. (1990). Hotelling's theorem on the volume of tubes: some illustrations in simultaneous inference and data analysis. *Ann. Statist.* **18** 652–684.

KNAFL, G., SACKS, J. and YLVISAKER, D. (1985). Confidence bands for regression functions. *J. Amer. Statist. Assoc.* **80** 683–691.

KNOWLES, M. and SIEGMUND, D. (1989). On Hotelling's geometric approach to testing for a nonlinear parameter in regression. *Internat. Statist. Rev.* **57** 205–220.

KREYSZIG, E. (1968). *Introduction to Differential Geometry and Riemannian Geometry*. Univ. Toronto Press.

LOADER, C. (1993). Nonparametric regression, confidence bands and bias correction. In *Computing Science and Statistics. Proceedings of the 25th Symposium on the Interface* 131–136. Interface Foundation of North America, Fairfax Station, VA.

LOADER, C. and SUN, J. (1997). Robustness of tube formula. *J. Comput. Graph. Statist.* **6** 242–250.

NAIMAN, D. Q. (1987). Simultaneous confidence bounds in multiple regression using predictor variable constraints. *J. Amer. Statist. Assoc.* **82** 214–219.

NAIMAN, D. Q. (1990). On volumes of tubular neighborhoods of spherical polyhedra and statistical inference. *Ann. Statist.* **18** 685–716.

NAIMAN, D. Q. and WYNN, H. P. (1997). Abstract tubes, improved inclusion–exclusion identities and inequalities and importance sampling. *Ann. Statist.* **25** 1954–1983.

SCHEFFÉ, H. (1959). *The Analysis of Variance*. Wiley, New York.

SKOVGAARD, I. M. (1981a). Transformation of an Edgeworth expansion by a sequence of smooth functions. *Scand. J. Statist.* **8** 207–217.

SKOVGAARD, I. M. (1981b). Edgeworth expansions of the distribution of maximum likelihood estimators in the general (non i.i.d.) case. *Scand. J. Statist.* **8** 227–236.

SUN, J. (1993). Tail probabilities of the maxima of Gaussian random fields. *Ann. Probab.* **21** 34–71.

SUN, J. and LOADER, C. (1994). Simultaneous confidence bands for linear regression and smoothing. *Ann. Statist.* **22** 1328–1345.

SUN, J., LOADER, C. and MCCORMICK, W. (1998). Confidence bands for generalized confidence bands. Available at http://sun.cwru.edu/˜jiayang/glm.ps.

SUN, J., RAZ., J. and FARAWAY, J. (1999). Simultaneous confidence bands for growth and response curves. *Statist. Sinica* **9** 679–698.

WEYL, H. (1939). On the volume of tubes. *Amer. J. Math.* **61** 461–472.

J. Sun
Department of Statistics
Case Western Reserve University
Cleveland, OH 44108-7054
E-mail: jiayang@sun.cwru.edu

C. Loader
Bell Laboratories
Lucent Technologies
700 Mountain Avenue
Murray Hill, NJ 07974–0636

W. P. McCormick
Department of Statistics
University of Georgia
Athens, GA 30602–1952