The Annals of Statistics
1986, Vol. 14, No. 1, 326–335

BEHAVIOUR OF THIRD ORDER TERMS IN QUADRATIC APPROXIMATIONS OF LR-STATISTICS IN MULTIVARIATE GENERALIZED LINEAR MODELS

BY CHRISTIAN KREDLER

Technical University, Munich

The size of error is investigated when the log-likelihood of multivariate generalized linear models is approximated by a quadratic function. The nonquadratic tail is characterized by analyzing the cubic part of the log-likelihood. In a local analysis simple bounds for that part can be expressed in terms of expectations of the related random variables for arbitrary sample size N. Additionally global error bounds are given for the univariate case.

Introduction. In many applications the log-likelihood is approximated by a quadratic function. An adequate approximation and the corresponding asymptotic theory ensure that the classical linear regression results apply to nonnormal models as well; cf. Lawless and Singhal (1978). The adequacy depends on the parametrization (cf. Section 2) and on the size of the global approximation error, for which until now no bounds have been published.

However, the investigations of Minkin (1983) and of this paper may be viewed as steps in this direction. Both articles assess local errors in the main. Minkin (1983) considers independent observations $y_{(i)}$ of a univariate one parameter family and the linear relationship $\theta_{(i)} = x_{(i)}^T\beta$. He gives relative and absolute bounds for the approximation error in a region R_a characterized by a quadratic form using the information matrix

$$Q(\beta, \hat{\beta}) = (\beta - \hat{\beta})^T J\{\hat{\beta}\}(\beta - \hat{\beta}),$$

namely

$$R_a = \{\beta | Q(\beta, \hat{\beta}) \leq a\}.$$

Minkin's bounds depend on the analysis of more or less complicated remainder functions. The paper treats logit and log-linear models explicitly, which are special cases of univariate generalized linear models (GLM's); cf. Nelder and Wedderburn (1972).

In this article global error bounds for logit, log-linear, and models depending on the gamma distribution are presented. The main subject, however, is a local analysis of the relative quadratic approximation error in multivariate canonical GLM's. Hence we cover the important multinomial case.

In Section 1 the third derivative structure of an exponential family with canonical parametrization is analyzed. A general technique for obtaining bounds, applicable to all distributions of this kind, is derived in Section 2 and explicit

Received November 1983; revised August 1985.
AMS 1980 *subject classifications.* Primary 62E10; secondary 62F04, 62F07, 62J02.
Key words and phrases. Exponential families, third moments, generalized linear models, LR-statistic, quadratic approximation error, variable selection.

results for the Poisson, gamma, and multinomial distribution follow. In any case it turns out that the size of the cubic terms compared with that of the quadratic one is of the same order as the relative deviations in the expectations of the related random variables.

In the general multivariate case only local error bounds are available. Nevertheless they reveal what really happens in the nonquadratic tail. In Section 3 this can be seen by comparing global and local errors for the univariate distributions mentioned above. The results can be used in Section 4 to obtain a bound for the cubic error terms in the quadratic approximation of likelihood ratio (LR-statistics in GLM's which is valid for arbitrary sample size N.

Finally it is shown how the results can be used to test linear hypotheses in common GLM applications, like log-linear models or contingency tables. A special multinomial regression model with continuous factors, namely the logistic discriminant approach [cf. Anderson (1972)], completes the last section.

1. Third order properties of exponential family densities. Before we can treat the subject of our paper we summarize some properties of exponential families in canonical parametrization and the multinomial distribution which are needed later.

Consider a discrete or continuous random variable $Y = (Y_j) \in \mathbb{R}^q$ having an exponential family density

$$f(y|\theta) = c(y)\exp\{y^T\theta - b(\theta)\} \tag{1}$$

with respect to the counting or Lebesgue measure. The canonical parameter $\theta = (\theta_j) \in \mathbb{R}^q$ is not involved in c. The partial derivatives of b, b_j, b_{jk}, $b_{jkl}, \dots$ generate the cumulants of Y. With

$$\begin{aligned} \mu &= (\mu_j) = E(Y) = (b_j), \\ \Sigma &= (\sigma_{jk}) = \operatorname{cov}(Y_j, Y_k) = (b_{jk}), \quad \text{etc.} \end{aligned} \tag{2}$$

According to Barndorff-Nielsen (1978) the correspondence between θ and μ is one to one and there is a scalar function $g(\mu)$, namely the Legendre transform of b, with partial derivatives $g_r, g_{rs}, g_{rst}, \dots$ such that

$$\begin{aligned} \theta_r &= g_r(\mu), \\ \Sigma^{-1} &= (g_{rs}(\mu)), \end{aligned} \tag{3}$$

and

$$b_{jkl} = -\sum_{r,s,t} g_{rst}(\mu)\sigma_{rj}\sigma_{sk}\sigma_{tl}, \qquad j,k,l,r,s,t = 1,\dots,q. \tag{4}$$

The partial derivatives of $g(\cdot)$ appear in related calculations by McCullagh (1984). If we regard the third derivative tensors $\mathbf{B} = (b_{jkl})$, $\mathbf{G} = (g_{rst})$ as operators with three arbitrary arguments u, v, w and $\bar{u} = \Sigma u$, $\bar{v} = \Sigma v$, and $\bar{w} = \Sigma w \in \mathbb{R}^q$, respectively, we obtain from (4)

$$\begin{aligned} \mathbf{B}[u,v,w] &= \sum_{j,k,l} b_{jkl}u_j v_k w_l = -\sum_{r,s,t} g_{rst}\bar{u}_r\bar{v}_s\bar{w}_t \\ &= -\mathbf{G}[\Sigma u, \Sigma v, \Sigma w]. \end{aligned} \tag{5}$$

Finally we give the expressions above explicitly for the multinomial case, because these relations are needed in the proofs of the subsequent section. There the factorization (5) and the simple structure of **G** play a fundamental role.

Since the multinomial $\bar{Y} = (Y_0, Y_1, \dots, Y_q)^T$ with $E(\bar{Y}) = (\pi_0, \pi_1, \dots, \pi_q)^T$, $\pi_0 + \cdots + \pi_q = 1$ has a singular distribution, Y_0 does not appear in (1). We obtain $b(\theta) = \ln(1 + e^{\theta_1} + \cdots + e^{\theta_q})$ and with the Kronecker delta δ_{jk},

$$\Sigma^{-1} = (g_{rs}) = \left(\frac{1}{\pi_0} + \frac{\delta_{rs}}{\pi_r} \right), \tag{6}$$

$$\mathbf{G} = (g_{rst}) = \left(\frac{1}{\pi_0^2} - \frac{\delta_{rs}\delta_{st}}{\pi_r^2} \right), \qquad r, s, t = 1, \dots, q, \tag{7}$$

and hence with $\bar{u}, \bar{v}, \bar{w}$ from (5)

$$\mathbf{B}[u, v, w] = \bar{u}^T(\tau_{rs}\{\pi, \bar{w}\})\bar{v}, \tag{8}$$

where

$$\tau_{rs} = \delta_{rs}\frac{\bar{w}_r}{\pi_r^2} - \sum_t \bar{w}_t/\pi_0^2 \quad \text{and} \quad \pi = (\pi_j).$$

2. Local error bounds in quadratic log-likelihood approximation. For normally distributed random variables with known variance the log-likelihood is quadratic in the mean value parameter. Outside of this special case

$$L(\theta) := \ln f(y|\theta) \tag{9}$$

is often approximated by some quadratic function. The adequacy depends heavily on the parametrization used. Here we investigate quadratic Taylor polynomials for the canonical parameter θ. In this case derivatives of order greater than 1 do not involve the random variable y. Hence the bounds given in Section 4 for GLM's are valid for every sample size N (not only asymptotically). Here we present our results concerning the structure of the cubic terms in exponential family log-likelihoods. Global bounds are to be found in the subsequent section.

$L(\theta)$ is expanded about some fixed θ^* and μ, μ^* denote the corresponding expectations. It turns out that relative errors in the components of μ and μ^* are responsible for an inadequate quadratic approximation, i.e.,

$$\epsilon = O\left(\max_j \left| \frac{\mu_j - \mu_j^*}{\mu_j^*} \right| \right).$$

With the quadratic form

$$Q(\theta, \theta^*) := (\theta - \theta^*)^T \Sigma\{\theta^*\}(\theta - \theta^*), \tag{10}$$

the Taylor expansion of $L(\theta)$ is

$$\begin{aligned} L(\theta) - L(\theta^*) = (y - \mu^*)^T(\theta - \theta^*) - \tfrac{1}{2}Q(\theta, \theta^*)\left(1 + \tfrac{1}{3}\rho_3(\theta, \theta^*)\right) \\ + R_4(\theta, \theta^*), \end{aligned} \tag{11}$$

where $-\rho_3 Q/6$ and R_4 describe the terms of order 3 and 4, respectively.

2.1. The univariate case. For simplicity we first present the results for univariate Y, i.e., $q = 1$. Denote by $b^{(j)}$ and $g^{(j)}$ the j-order derivatives of b and g, respectively (e.g., $g^{(3)} = g_{111}$), and let $\sigma_*^2 = b^{(2)}(\theta^*) = \sigma_{11}(\theta^*)$. Now (11) can be written as

$$(12)\qquad \begin{aligned} L(\theta) - L(\theta^*) = (\mu - \mu^*)(\theta - \theta^*) - \tfrac{1}{2}\sigma_*^2(\theta - \theta^*)^2 \\ \times\left[1 + \tfrac{1}{3}b^{(3)}(\theta^*)(\theta - \theta^*)/\sigma_*^2\right] + R_4(\theta, \theta^*). \end{aligned}$$

From the results of Section 1 we deduce

$$(13)\qquad \theta - \theta^* = \frac{(\mu - \mu^*)}{\sigma_*^2} + O(\mu - \mu^*)^2.$$

Together with (5) this shows that ρ_3 and

$$(14)\qquad \rho_3^* = -g^{(3)}(\mu^*)\sigma_*^2(\mu - \mu^*)$$

differ only by terms of order 2. Hence we can replace ρ_3 by ρ_3^* in (11) and get a new remainder R_4^* which is of the same order as the original R_4 (as ρ_3 is multiplied by Q). Using the notation of (3) we obtain

$$(15)\qquad \theta = \ln(\mu) = g_1(\mu); \qquad g^{(3)}(\mu) = -\mu^{-2}; \qquad \sigma^2 = \mu = b^{(2)}(\theta)$$

for the Poisson and for $\Gamma(\nu, \gamma)$ gamma variables:

$$(16)\qquad \begin{aligned} \theta = -\gamma = -\nu/\mu = g_1(\mu); \\ g^{(3)}(\mu) = -2\nu\mu^{-3}; \qquad \sigma^2 = \mu^2/\nu = b^{(2)}(\theta). \end{aligned}$$

This leads to our first crucial result:

THEOREM 2.1. *Size of ρ_3^* for Poisson and gamma distribution. The relative size of the cubic terms compared with the quadratic ones in* (11) *is*

$$(17)\qquad |\rho_3^*| = C\frac{|\mu - \mu^*|}{\mu^*},$$

where $C_{\text{Poisson}} = 1$, $C_{\text{gamma}} = 2$.

2.2. The multinomial case. In the multivariate case ρ_3 of (11) can be replaced again by ρ_3^*. After this we can formulate the following theorem for the quadratic approximation error of the log-likelihood:

THEOREM 2.2. *Size of ρ_3^* in the binomial and multinomial case. Let $\pi_0, \dots, \pi_q$ as given in Section* 1; *then for $q = 1$ (binomial) we obtain*

$$(18)\qquad \rho_3^* = \frac{\pi_1 - \pi_1^*}{\pi_1^*} + \frac{\pi_0 - \pi_0^*}{\pi_0^*},$$

whereas for $q > 1$

$$(19)\qquad |\rho_3^*| \le 2\sqrt{(q + 1)} \max_{0 \le m \le q} \frac{|\pi_m - \pi_m^*|}{\pi_m^*}.$$

The bound in (18) *is slightly stronger than* (19) *for* $q = 1$. *The additional factor* $(q+1)^{1/2}$ *is due to Lemma* 2.5 (*formulation to follow*).

Before proving (18)–(19) we expose how these results can be used in practical applications.

REMARK 2.3. *Size of quadratic approximation error.* From the theorems we conclude that the relative size of the cubic terms is small if the leading digits of the expectations μ and μ^* coincide. In Section 3 we show that small ρ_3 yields a small global approximation error R_4 under very general conditions. Hence the approximation error is given in the main by the simple formulas (17)–(19). Conversely, if the leading digits of μ and μ^* differ, the quadratic approximation may be bad. So we have derived a simple criterion for the adequacy of quadratic log-likelihood approximations of exponential families (and as we see in Section 4, of GLM's), which can be easily checked during a computational procedure.

PROOF OF THEOREM 2.2. For the special binomial case we obtain with (7) and $\pi_0 = 1 - \pi_1$

$$\rho_3^* = (\pi_1 - \pi_1^*)\frac{1 - 2\pi_1^*}{\pi_1^*\pi_0^*} = (\pi_1 - \pi_1^*)\left(\frac{1}{\pi_1^*} - \frac{1}{\pi_0^*}\right), \tag{20}$$

which is the same as (18). □

Since the proof for multinomial Y is tedious only the main steps are sketched.

OUTLINE OF THE PROOF FOR MULTINOMIAL Y. Let $\Sigma := \Sigma\{\theta^*\}$ and $T\{\cdot\} = (\tau_{rs}\{\pi^*, \cdot\})$ which are positive definite or symmetric, respectively, and further $\pi = (\pi_1, \dots, \pi_q)^T$ and $e = (1, \dots, 1)^T$. We need two linear algebra lemmas which are stated without proof:

LEMMA 2.4. *Factorization of the spectral radius* r *of a symmetric product. For arbitrary* $u \in \mathbb{R}^q$

$$u^T\Sigma T\Sigma u \le r\left[\Sigma^{1/2}T\Sigma^{1/2}\right]u^T\Sigma u. \tag{21}$$

LEMMA 2.5. *Bounds for spectral norms.*

$$\begin{aligned} \|I - \pi e^T\|_2 &\le \sqrt{q+1}, \\ \|\pi e^T\|_2 &\le \sqrt{q}. \end{aligned} \tag{22}$$

Now the three steps of the multivariate proof are as follows:

(i) Let $u = v = w := \theta - \theta^*$. Then (8) and Lemma 2.4 yield:

$$\begin{aligned} |\mathbf{B}[w, w, w]| = |w^T\Sigma T\{\Sigma w\}\Sigma w| &\le r\left[\Sigma^{1/2}T\{\Sigma w\}\Sigma^{1/2}\right]Q(\theta, \theta^*) \\ &= r\left[\Sigma T\{\Sigma(\theta - \theta^*)\}\right]Q(\theta, \theta^*). \end{aligned} \tag{23}$$

(ii) Analogously to (13) we expand $w = \theta - \theta^*$ and obtain

$$\Sigma(\theta - \theta^*) = \pi - \pi^* + \overline{R}_4, \tag{24}$$

where $\overline{R}_4$ again is multiplied by Q and hence is of order 4.

(iii) Define $p_m = (\pi_m - \pi_m^*)/\pi_m^*$, $m = 0, \dots, q$; $D\{p\} = \operatorname{diag}\{p_1, \dots, p_q\}$; then (6) and (8) yield in matrix notation

$$\begin{aligned} r[\Sigma T\{\pi - \pi^*\}] &\le \|(I - \pi e^T) D\{p\}\|_2 + \|\pi e^T\|_2 p_0 \\ &\le 2\sqrt{q+1} \max_m |p_m|, \end{aligned}$$

where we have used $r[A] \le \|A\|_2$ and applied Lemma 2.5. □

REMARK 2.6. The techniques described can be used in other cases as well, because steps (i) and (ii) work for any exponential family. Step (iii) and Lemma 2.5 apply to the class of "multinomial like" distributions with a Σ- and $\mathbf{G}$-structure analogous to (6)–(8), e.g., to the negative multinomial and the logarithmic series distribution [cf. Johnson and Kotz (1969)].

3. Global error bounds depending on the size of ρ_3. In this section the connection between the global approximation error R_4 in (12) and the relative error ρ_3 of cubic terms compared with quadratic terms is analyzed. For the Poisson and gamma distribution small ρ_3 causes a small R_4. This is not true for the binomial case. Here $\rho_3 = 0$ is possible, whereas the global quadratic approximation error can be quite large (cf. Section 3.2). Note that the bounds of the previous section hold for ρ_3^*. But according to (14), ρ_3 and ρ_3^* differ only by terms of order 2.

3.1. The principle. We use the notation of Section 2.1. For an arbitrary univariate exponential family the following approach works.

Starting from equation

$$\rho_3 = (\theta - \theta^*) b^{(3)}(\theta^*)/b^{(2)}(\theta^*) \tag{25}$$

the error term of (12) is

$$R_4(\theta, \theta^*) = \frac{1}{4!} b^{(4)}(\bar{\theta}) \rho_3^4 \left[\frac{b^{(2)}(\theta^*)}{b^{(3)}(\theta^*)}\right]^4, \tag{26}$$

where $\bar{\theta}$ is some convex combination of θ and θ^*.

Hence we get the following:

THEOREM 3.1 (Global approximation error). *The remainder R_4 of the Taylor expansion* (12) *can be bounded by*

$$|R_4| \le \rho_3^4 M/4!,$$

where

$$M_{\text{Poisson}} = \max\{e^{\theta}, e^{\theta^*}\}, \tag{27}$$

$$M_{\text{gamma}} = \frac{3\nu}{8}\max\left\{1, \frac{\theta^*}{\theta}\right\}, \tag{28}$$

$$M_{\text{binomial}} = \frac{(1+e^{\theta^*})^4}{(1-e^{\theta^*})^4}, \qquad \theta^* \neq 0. \tag{29}$$

PROOF. By inserting the following quantities

	$b^{(2)}(\theta^*)/b^{(3)}(\theta^*)$	$b^{(4)}(\bar{\theta})$
Poisson	1	$e^{\bar{\theta}}$
gamma	$-\theta^*/2$	$6\nu/\bar{\theta}^4$
binomial	$(1+e^{\theta^*})/(1-e^{\theta^*})$	$\lvert b^{(4)}(\bar{\theta})\rvert \le 1$ cf. Section 3.2

(27) and (28) are valid for arbitrary θ, θ^*, whereas the binomial bound approaches infinity for $\theta^* \to 0$. This case is investigated in the following section. □

3.2. Global bounds for the binomial distribution. The second order Taylor expansion of the binomial log-likelihood is

$$L(\theta) - L(\theta^*) = (y-\mu)(\theta-\theta^*) - \tfrac{1}{2}\sigma_*^2(\theta-\theta^*)^2 + R(\theta,\theta^*), \tag{30}$$

where $\sigma_*^2 = e^{\theta^*}/(1+e^{\theta^*})^2$. Since $b^{(3)}(\theta) = e^{\theta}(1-e^{\theta})/(1+e^{\theta})^3$ and $b^{(4)}(\theta) = e^{\theta}(1-4e^{\theta}+e^{2\theta})/(1+e^{\theta})^4$ are uniformly bounded by 1 we have:

THEOREM 3.2 (Global quadratic approximation error of the binomial). *The remainder of the log-likelihood Taylor expansion* (30) *is bounded by*

$$|R(\theta,\theta^*)| \le |\theta-\theta^*|^3/3! \tag{31}$$

and for $\theta^* = 0$ *we obtain*

$$|R(\theta,\theta^*)| \le \theta^4/4!. \tag{32}$$

3.3. Interpretation of the bounds. Working with quadratic log-likelihood approximation we are interested in how large the neglected terms are compared with the approximating ones. This problem is solved exactly only for univariate Y.

Except the binomial case for $\rho_3 = 0$, which has been analyzed separately, for Poisson, gamma, and binomial random variables small ρ_3 yields a small global approximation error. This justifies the practical relevance of ρ_3 as a measure for the adequacy of the quadratic approximation. Note that the roles of θ and θ^* in (11) and (12) may be interchanged which possibly improves the bounds (27)–(29), e.g., $M_{\text{gamma}} = \frac{3}{8}\nu$.

An explicit analysis for the multinomial case seems too hard because of the difficult multivariate structure.

In the last chapter we demonstrate how the results can be applied to model selection in canonical generalized linear models.

4. LR-statistic approximation and model selection in GLM's. The theorems of Section 2 refer to the log-likelihood L itself and not to any expectation of L or its derivatives. In the following the results are extended in a natural way to generalized linear models (GLM's). Hence we obtain computationally available bounds for the quadratic approximation error of the according LR-statistic which are valid for any sample size N. An application of hypothesis testing in logistic discriminant analysis completes this section.

Consider independent exponential family response variables $y_{(i)} \in \mathbb{R}^q$ characterized by parameters $\theta_{(i)} \in \mathbb{R}^q$ and linked with regressor variables $x_{(i)} \in \mathbb{R}^P$, $i = 1, \dots, N$. According to (1) the common log-likelihood (here denoted by L_N) is given by

$$L_N(\vec{\theta}) = \text{const.} + \sum_i \left\{ y_{(i)}^T \theta_{(i)} - b\left(\theta_{(i)}\right) \right\}, \tag{33}$$

where $\vec{\theta}$ is the collection of the $\theta_{(i)}$.

Although our approach covers the multinomial case, too, we restrict ourselves to $q = 1$, for simplicity, and consider GLM's, where the unknown parameters $\theta_{(i)} \in \mathbb{R}$ range in a linear subspace, spanned by the $x_{(i)}$, i.e.,

$$\vec{\theta} = X\beta. \tag{34}$$

$X = (x_{(1)}^T, \dots, x_{(N)}^T)^T$ denotes the usual design matrix and $\beta \in \mathbb{R}^P$ is an unknown parameter vector.

In the notation of Nelder and Wedderburn (1972) the model described by (33) and (34) is a canonical GLM, because the $\theta_{(i)}$ themselves, and not some functions of them, range in a linear subspace.

Among several authors McCullagh and Nelder (1983) expose that a variety of classical applications, like log-linear models, contingency tables, and binary and survival data analysis fit into the GLM framework. Furthermore, according to Fahrmeir and Kredler (1984), the logistic discrimination approach, which seems to be superior to the widely used linear Fisher discrimination, [cf. Press and Wilson (1978)], leads formally to a multinomial GLM.

Model selection and checking the goodness of fit in the models mentioned above usually is carried out by tests of hypotheses H, concerning the parameters $\vec{\theta}$ or β, respectively. Denote by $\hat{\theta}_H = X\hat{\beta}_H$ and $\hat{\theta} = X\hat{\beta}$ the maximizer of L_N under and without restriction H. As $\nabla L_N(\hat{\theta}) = 0$ the Taylor expansion of the LR-statistic is

$$\begin{aligned} \lambda &= -2\left\{ L_N(\hat{\theta}_H) - L_N(\hat{\theta}) \right\} \\ &= Q_N(\hat{\theta}_H, \hat{\theta})(1 + \rho_N^*/3) - 2R_N^{*(4)}, \end{aligned} \tag{35}$$

where Q_N, ρ_n^*, and $R_N^{*(4)}$ denote the quantities (associated with the total sample size N) corresponding to Q, ρ_3^*, and R_4^* of Section 2.

Applying the theorems of Section 2 yields directly:

THEOREM 4.1 (Size of error in quadratic LR-statistic approximation). *For Poisson, gamma, and multinomial response variables we obtain*

$$|\rho_N^*| \leq 2\sqrt{q+1} \max_i \max_{0 \leq m \leq q} \left| \frac{\mu_{im}^H - \hat{\mu}_{im}}{\hat{\mu}_{im}} \right|, \tag{36}$$

where $\hat{\mu}_{ij}(\hat{\mu}_{i0} = 1 - \hat{\mu}_{i1} - \cdots - \hat{\mu}_{iq}$ *for multinomial* $y_{(i)})$ *and* μ_{ij}^H *denote the estimated expectations under and without restriction* H. *According to* (36) *the* LR-*statistic can be replaced by its quadratic Taylor polynomial; presumably the relative errors of* $\hat{\mu}_{ij}$ *and* μ_{ij}^H *are small compared with* 1, *i.e., the leading figures coincide. For the test of linear hypotheses*

$$H: A\beta = a$$

in the models mentioned above we can proceed as follows:

(*i*) *Compute* $\hat{\theta}$, $L_N(\hat{\theta})$, *and* $-\nabla^2 L_N(\hat{\theta})$.

(*ii*) *Compute approximations for* $\hat{\beta}_H$, $\hat{\theta}_H$, *and the LR-statistic* λ *in a suitable linear model with cross-product matrix* $-\nabla^2 L_N(\hat{\theta})$.

(*iii*) *Accept the quadratic approximations of* (*ii*) *if the leading digits of* $\hat{\mu}_{ij}$ *and* μ_{ij}^H *coincide.*

Usually during the process of model building a variety of nested linear hypotheses, e.g., concerning the significance of certain parameters, are carried out. For a sequence of tests the same linear model can be used as long as condition (iii) is fulfilled. If (iii) is violated we take a new adjusted cross-product matrix instead of $-\nabla^2 L_N(\hat{\theta})$. For example, in logistic discriminant analysis [cf. Anderson (1972)], the a posteriori probability for the classification of observation (i) in group Ω_k, $k = 0, 1, \ldots, q$ is given by

$$\pi(x_i, \beta_{(k)}) = \exp(x_i^T, \beta_{(k)}) \Big/ \sum_{m=0}^{q} \exp(x_i^T \beta_{(m)}), \qquad \beta_{(0)} \equiv 0.$$

This leads to a canonical multinomial GLM. We set $x_{i1} \equiv 1$ and start with the explanatory variables $x_{i2}, \ldots, x_{ip}$. In order to check which variables are significant for discrimination, we test nested hypotheses like

$$H_j: \beta_p = \beta_{p-1} = \cdots = \beta_j = 0, \qquad j = p, \ldots, 2.$$

The corresponding values of the log-likelihood $L_N(\hat{\theta}_{H_j})$, then can be computed in the following way:

$$L_N(\hat{\theta})\{\text{exact}\}, L_N(\hat{\theta}_{H_p}), \ldots, L_N(\hat{\theta}_{H_{k+1}})\{\text{quadratic approximation}\}$$

$$\text{update } L_N(\hat{\theta}_{H_k})\{\text{exact}\}, L_N(\hat{\theta}_{H_{k-1}}), \ldots \{\text{quadratic approximation}\}.$$

For variable selection problems this yields an effective and fast procedure [cf. Kredler (1984)].

5. Conclusions. The analysis of the LR-statistic for multinomial, Poisson, and gamma samples, given in this paper, applies in an analogous way to related distributions like the negative multinomial and logarithmic series. This paper gives an algebraic analysis of the quadratic log-likelihood approximation. The analysis depends heavily on the use of the canonical parametrization of exponential families, only in this case higher derivatives of the log-likelihood do not depend on the random variable Y. Consequently the error bounds for canonical GLM's given in the last section are valid for arbitrary sample size (not only asymptotically). Since the analytic dependencies between canonical and expectation parameters play a crucial role our results cannot be easily extended to other parametrizations. Nevertheless our analysis works for many important applications.

Acknowledgments. I would like to thank the referees and the editor for helpful comments and constructive suggestions on earlier versions of this paper.

REFERENCES

ANDERSON, J. A. (1972). Separate sample logistic discrimination. *Biometrika* **59** 19–35.

BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families*. Wiley, Chichester.

FAHRMEIR, L. and KREDLER, C. (1984). Verallgemeinerte lineare Modelle. In *Multivariate Statistische Verfahren*. (L. Fahrmeir, and A. Hamerle, eds.) De Gruyter, Berlin.

JOHNSON, N. L. and KOTZ, S. (1969). *Discrete Distributions*. Wiley, New York.

KREDLER, C. (1984). Selection of variables in certain nonlinear regression models. *Computational Statist. Quart.* **1** 13–27.

LAWLESS, J. F. and SINGHAL, K. (1978). Efficient screening of nonnormal regression models. *Biometrics* **34** 318–327.

MCCULLAGH, P. (1984). Tensor notation and cumulants of polynomials. *Biometrika* **71** 461–476.

MCCULLAGH, P. and NELDER, J. A. (1983). *Generalized Linear Models*. Chapman and Hall, London.

MINKIN, S. (1983). Assessing the quadratic approximation to the log likelihood function in nonnormal linear models. *Biometrika* **70** 367–372.

NELDER, J. A. and WEDDERBURN, R. W. M. (1972). Generalized linear models. *J. Roy. Statist. Soc. Ser. A* **135** 370–384.

PRESS, S. J. and WILSON, S. (1978). Choosing between logistic regression and discriminant analysis. *J. Amer. Statist. Assoc.* **73** 699–705.

INSTITUT FÜR ANGEWANDTE
MATHEMATIK UND STATISTIK
TU MÜNCHEN
ARCISSTRAßE 21
8000 MÜNCHEN 2
WEST GERMANY