Statistische Hefte 27, 179–205 (1986)

Statistische Hefte
Statistical Papers

Asymptotic inference in discrete response models

Ludwig Fahrmeir and Heinz Kaufmann

Received: March 15, 86; Revised version: August 19, 86

Regression models for discrete responses have found numerous applications. We consider logit, probit and cumulative logit models for qualitative data, and the loglinear and linear Poisson model for counted data. Statistical analysis of these models relies heavily on asvmptotic likelihood theory, i.e. asymptotic properties of the maximum likelihood estimator and the likelihood ratio as well as related test statistics. In practical situations, previously published conditions assuring these properties may be too strong, or it is difficult to see whether they apply. This paper contributes to a clarification of this point and characterizes to some extent situations where asymptotic theory is applicable and where it is not. In particular, sharp upper bounds on the admissible growth of regressors are given.

Keywords: qualitative response models, Poisson models, asymptotic inference, maximum likelihood, likelihood ratio.

1. INTRODUCTION. In recent years, regression models for discrete responses have found numerous applications, and a number of programming packages make it possible now to fit such models. Statistical inference relies heavily on asymptotic likelihood theory. Conditions to assure asymptotic properties of the maximum likelihood estimator (MLE) of regression coefficients have previously been given by Haberman (1977a,b), Nordberg (1980), Gourieroux and Monfort (1981), Fahrmeir and Kaufmann (1985, henceforth abbreviated by F/K). Stochastically dependent polytomous responses are considered by Kaufmann (1987), and the likelihood ratio and related statistics are investigated by Mathieu (1981) and Fahrmeir (1986).

In practical situations, the researcher is confronted with particular sequences of covariates and has to decide whether asymptotic claims hold. However, conditions given by the authors above may be too strong, or they are in such a general form that it is difficult to see whether they apply. This paper contributes, for the most common models, to a clarification of this point and characterizes to some extent situations where asymptotic theory is applicable and where it is not.

The models treated are shortly described in Section 2 (for a more detailed exposition and further references, see e.g. McCullagh and Nelder, 1983, Fahrmeir and Hamerle, 1984, ch.6,7 and 10). We consider the logit and probit model for binary responses, the multinomial resp. cumulative logit model for unordered resp. ordered multinomial responses, and the loglinear and linear Poisson model for counted data.

For all these models we assume that the data form a sequence $\{(y_n,z_n),n \in \mathbb{N}\}$ resp. $\{(y_n,x_n),n \in \mathbb{N}\}$ of stochastically independent observations, where y_n denotes the n-th response and z_n resp. x_n the corresponding vector of regressors or covariates. The symbol x_n is reserved for covariates without a constant term; more precisely, there must not exist a λ such that $x_n'\lambda = 1$ for all $n \in \mathbb{N}$. Commonly, one specifies $z_n = (1,x_n')'$ allowing for a constant. Further, the covariates may be random or not. If they are random, it is supposed that the pairs $\{(y_n,z_n)\}$ resp. $\{(y_n,x_n)\}$ are i.i.d. as a pair (y,z) resp. (y,x) of random variables.

The density of y_n given the covariates z_n resp. x_n is assumed to depend on linear combinations of these covariates. The unknown coefficients of these linear combinations are collected together in the p-vector β, say. Maximum likelihood estimation of β as well as tests of linear hypotheses on β are discussed shortly in Section 3.

Asymptotic results are given in Section 4. Theorem 1 states asymptotic existence, consistency and asymptotic normality of the MLE. Due to concavity of the log likelihood, claims are more definite than in F/K. Theorems on asymptotic efficiency of the MLE and asymptotic χ^2-distribution of the likelihood ratio and related statistics are added.

In Subsections 4.1 to 4.4, conditions are given under which the assertions of these theorems hold. In Subsection 4.1, bounded regressors are treated. Then the divergence condition

$$\lambda_{min} \sum_{i=1}^{n} z_i z_i' \to \infty , \tag{1.1}$$

which is well known from classical linear regression, is sufficient (the symbol λ_{min} denotes smallest eigenvalue).

In Subsection (4.2) growing regressors are allowed. Under a condition somewhat sharper than (1.1) the admissible growth rate is

$$\|z_n\| = o(\log n) ,$$

for all models with the exception of the probit and linear Poisson model, where the admissible rates are

$\|z_n\|^2 = o(\log n)$ resp. $\|z_n\| = O(n^{\alpha})$, with some $\alpha > 0$.

Partial converses demonstrate that these rates are sharp.

Stochastic regressors are treated in Subsection 4.3. Here the assertions hold under a condition assuring essentially identifiability of β, and some moment condition on the marginal distribution of z resp. x differing from model to model.

Conditions of Subsections 4.1-4.3 involve only asymptotic behaviour of $\{z_n\}$, resp. tail behaviour of z for stochastic regressors. In Subsection 4.4 rather general conditions are given which might be useful for situations not covered by 4.1 to 4.3. Some finite sample insight is given by the final remark of 4.4, as well as by the simulation results of Section 5.

Proofs are given in Section 6. Mainly, conditions are shown to be sufficient for conditions of F/K and Fahrmeir (1986), using some lemmas which may be of interest in similar cases. Particularly we mention Lemma 2 stating conditions for multinomial observations.

2. MODELS. For a sequence $\{y_n\}$ of independent binary responses, with $y_n = 0$ ($y_n=1$) if the first resp. second category is observed at the n-th trial, define $\pi_n = P(y_n=1|z_n)$. We consider the *logit model*

$$\log(\pi_n/(1-\pi_n)) = z_n'\beta, \quad n \geq 1 ,$$

or equivalently

$$\pi_n = (1+\exp(-z_n'\beta))^{-1}, \quad n \geq 1 , \tag{2.1}$$

and the *probit model*

$$\pi_n = \phi\,(z_n'\beta), \quad n \geq 1 , \tag{2.2}$$

with the standard normal distribution function ϕ.

If each response can assume more than two, say m, categories, we set $y_{nj} = 1$, if the j-th category is observed at the n-th trial, and $y_{nj} = 0$, otherwise, $j=1,\ldots,m$, and assume that the n-th response is given as the vector $y_n = (y_{n1},\ldots,y_{nq})$ with $q = m-1$ components, the redundant component y_{nm} being dropped. Correspondingly, we set $\pi_{nj} = P(y_{nj} = 1|z_n)$, $j=1,\ldots,m$, and $\pi_n = (\pi_{n1},\ldots,\pi_{nq})'$. If the response categories are unordered, the most common model is the *multinomial logit model*

$$\log(\pi_{nj}/\pi_{nm}) = z_n'\beta_j , \quad j=1,\ldots,q, \quad n \geq 1 ,$$

or equivalently, with $v_n = (1+\sum_{j=1}^{q} \exp(z_n'\beta_j))^{-1}$,

$$\begin{aligned} \pi_{nj} &= v_n \exp(z_n'\beta_j), \; j=1,\ldots,q , \\ \pi_{nm} &= v_n , \qquad n \geq 1 . \end{aligned} \tag{2.3}$$

The parameter vector β is now given by $\beta' = (\beta_1',\ldots,\beta_q')$.

For ordered categories, we consider the *cumulative logit model*

$$\pi_{nj} = F(\delta_j - x_n'\lambda) - F(\delta_{j-1} - x_n'\lambda), \; j=1,\ldots,m, \quad n \geq 1, \tag{2.4}$$

with $F(\gamma) = (1+e^{-\gamma})^{-1}$. In order that these probabilities are strictly positive and sum up to one, we suppose that

$$-\infty = \delta_o < \delta_1 < \ldots < \delta_q < \delta_m = \infty . \tag{2.5}$$

Such a model may be motivated by assuming that, for the n-th trial, category j is observed if the value of an underlying latent variable falls into the j-th interval (δ_{j-1},δ_j), and the latent variable is the sum of its expectation $x_n'\lambda$

and a logistically distributed error term. Due to the introduction of $\delta_o,\dots,\delta_m$ the vectors x_n, $n \geq 1$, of regressors must not contain a constant term. For convenience, we set $z_n' = (1,x_n')$, $n \geq 1$. In estimating the vector $\beta = (\delta_1,\dots,\delta_q,\lambda')'$ restriction (2.5) has to be observed.

If $\{y_n\}$ is a sequence of independent counts, for given regressors, we assume that y_n is Poisson distributed with a strictly positive rate λ_n, say, and consider the *log linear Poisson model*

$$\log \lambda_n = z_n'\beta, \quad n \geq 1 , \tag{2.6}$$

and the *linear Poisson model*

$$\lambda_n = z_n'\beta, \quad n \geq 1 . \tag{2.7}$$

For the linear Poisson model, β must be restricted to

$$\{\beta, z_n'\beta > 0,\ n \in \mathbb{N}\} ,$$

to obtain strictly positive rates. A bit stronger, we restrict β to

$$B = \text{interior } \{\beta, z_n'\beta > 0, n \in \mathbb{N}\}, \tag{2.8}$$

since for boundary points additional asymptotic considerations are necessary. The admissible set B is assumed to be nonvoid (compare Lemma 3 in Section 6).

All the models above belong to the family of generalized linear models: the observations $\{y_n\}$ form a sequence of independent dichotomous, polytomous or Poisson responses, and

$$E\ y_n (= E_\beta y_n) = h(Z_n'\beta), \quad n \geq 1 , \tag{2.9}$$

where β is a p-vector of unknown parameters, out of an admissible set B, and Z_n is a p×q-matrix of regressors. For binary or Poisson responses, q=1 and $Z_n=z_n$. For (2.3) and (2.4), we have

$$Z_n' = \begin{bmatrix} z_n' & & & 0 \\ & z_n' & & \\ & & \ddots & \\ 0 & & & z_n' \end{bmatrix}, \quad Z_n' = \begin{bmatrix} 1 & & & -x_n' \\ & 1 & & -x_n' \\ & & \ddots & \vdots \\ & & 1 & -x_n' \end{bmatrix}, \tag{2.10}$$

respectively. The *link function* h can easily be found for each particular model.

With the exception of (2.4) and (2.7), where the restriction (2.5) resp. (2.8) has to be observed, the parameter set B is supposed to be the whole space $\mathbb{R}^p$. In any case the assumptions on B imply that it is nonvoid, open and convex.

3. STATISTICAL INFERENCE BASED ON MAXIMUM LIKELIHOOD ESTIMATION. Up to summands not depending on β, the log likelihood of the observations $y_1,..,y_n$ is

$$l_n(\beta) = \sum_{i=1}^{n} (y_i \log \pi_i + (1-y_i)\log(1-\pi_i)),\ \pi_i = h(z_i'\beta) ,$$

$$l_n(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{m} y_{ij} \log \pi_{ij},\ \pi_i = h(Z_i'\beta) ,$$

$$l_n(\beta) = \sum_{i=1}^{n}(y_i \log \lambda_i - \lambda_i),\ \lambda_i = h(z_i'\beta) ,$$

for dichotomous, polytomous resp. Poisson observations. If the regressors are stochastic, these are the log likelihoods given the observed regressors. The score function and the information matrix are

$$s_n(\beta) = \partial l \quad (\beta)/\partial\beta = \sum_{i=1}^{n} Z_i D_i(\beta)\Sigma_i^{-1}(\beta)(y_i-\mu_i(\beta)) ,$$

$$F_n(\beta) = \mathrm{cov}_\beta s_n(\beta) = \sum_{i=1}^{n} Z_i D_i(\beta)\Sigma_i^{-1}(\beta)D_i'(\beta)Z_i' ,$$

where $\mu_n(\beta)=E_\beta y_n=h(Z_n'\beta)$, $\Sigma_n(\beta)=\mathrm{cov}_\beta y_n$ and $D_n(\beta)=[\partial h(\gamma)/\partial\gamma]'$, evaluated at $Z_n'\beta$. The negative second derivative of the log likelihood will be denoted by $H_n(\beta)$.

For the models discussed in this paper, the log likelihood is concave, see Wedderburn (1976), Haberman (1980). Since B is open and convex, this implies that the zeros of the score function form a convex set maximizing the log likelihood globally. Hence we consider as MLE any random variable $\hat\beta_n$, depending on $y_1,\dots,y_n$, with $s_n(\hat\beta_n) = 0$, if $s_n(\beta), \beta \in B$, has a zero at all. If s_n has no zero within B, i.e. if no MLE exists, $\hat\beta_n$ may be defined as an arbitrary constant, to obtain a random variable defined throughout the sample space. Uniqueness of $\hat\beta_n$ is equivalent to full rank of $H_n(\hat\beta_n)$. For uniqueness conditions which can be checked without computing a MLE, see both authors above and Kaufmann (1986b,1986c). Finite sample conditions on the existence of a MLE are given by Wedderburn (1976), Haberman (1980), Albert and Anderson (1984), Silvapulle (1981). Asymptotically, the conditions of Section 4 imply existence of a MLE as well as its uniqueness. Computationally, a MLE can be obtained e.g. by the method of scoring.

Next let us consider a linear hypothesis

$$H_o\colon C\beta = \xi \quad \text{against} \quad H_1\colon C\beta \neq \xi , \tag{3.1}$$

where C has full row rank, say $r \leq p$, and the intersection $H_o \cap B$ is not empty. Then $H_o \cap B$ is a relatively open convex set of dimension p-r, and the log likelihood continues to be concave on $H_o \cap B$. Any unrestricted MLE will again be denoted by $\hat{\beta}_n$, whereas a MLE restricted to $H_o \cap B$ will be denoted by $\tilde{\beta}_n$. The likelihood ratio statistic

$$\lambda_n = -2\{l_n(\tilde{\beta}_n) - l_n(\hat{\beta}_n)\}$$

involves computation of both estimators. Due to concavity, it is unique, if it exists, i.e. if $\hat{\beta}_n$ and $\tilde{\beta}_n$ exist. Common approximations of λ_n are the Wald statistic

$$w_n = (C\hat{\beta}_n - \xi)'[CF_n^{-1}(\hat{\beta}_n)C']^{-1}(C\hat{\beta}_n - \xi)$$

and the score statistic

$$c_n = s_n'(\tilde{\beta}_n)F_n^{-1}(\tilde{\beta}_n)s_n(\tilde{\beta}_n) ,$$

requiring only computation of $\hat{\beta}_n$ resp. $\tilde{\beta}_n$.

4. ASYMPTOTIC THEORY. As norming quantities we use square roots of positive definite matrices, e.g. the Cholesky square root, see F/K.

Conditions to assure asymptotic properties, namely (B), (G), (S), (A), are stated in Subsections 4.1-4.4. The MLE is considered in Theorem 1 and Theorem 2.

THEOREM 1. Under (B), (G), (S) or (A), the probability that a unique MLE exists converges to one. Any sequence $\{\hat{\beta}_n\}$ of MLE's is consistent and asymptotically normal,

$$F_n^{T/2}(\beta)(\hat{\beta}_n-\beta) \xrightarrow{d} N(0,I) , \tag{4.1}$$

for any norming sequence $\{F_n^{T/2}(\beta)\}$ of square roots of the information matrix.

In conjunction with Theorem 1, several efficiency results can be derived, see Ibragimov and Hasminskii (1981,ch.II), Jeganathan (1982), Basawa and Scott (1983,Chapter 2), and Kaufmann (1987). Theorem 2 gives two of the possible efficiency results. The MLE $\hat{\beta}_n$ is compared with estimators $\tilde{\beta}_n$, which are functions of $y_1,\ldots,y_n$ and, for some sequence $\{F_n^{T/2}(\beta)\}$, regular in the following sense: for $\lambda \in \mathbb{R}^p$, set $\beta(n) = \beta + F_n^{-T/2}\lambda$. The estimator $\tilde{\beta}_n$, $n \geq 1$, is regular, if $\{F_n^{T/2}(\beta)(\tilde{\beta}_n - \beta(n))\}$ converges, under $P_{\beta(n)}$, in distribution to some random variable $V(\beta)$, for any fixed λ. For stochastic regressors, $P_{\beta(n)}$ denotes the conditional probability, given the regressors $\{Z_n\}$.

THEOREM 2. Under (B), (G), (S) or (A), the following statements hold for any norming sequence $\{F_n^{T/2}(\beta)\}$: any sequence $\{\hat{\beta}_n\}$ of MLE's is regular. Within the class of regular estimators $\{\tilde{\beta}_n\}$, the asymptotic probability of concentration,

$$\lim_{n\to\infty} P(F_n^{T/2}(\beta)(\tilde{\beta}_n-\beta)\in C)$$

attains its maximum if $\{\tilde{\beta}_n\} = \{\hat{\beta}_n\}$, for any symmetric convex set C. Within the class of regular estimators with a normally distributed limit vector $V(\beta) \sim N(0,\Sigma_\beta)$, say, the covariance matrix I of $\{\hat{\beta}_n\}$ is minimal in that the difference $\Sigma_\beta - I$ is always positive semidefinite.

For the testing problem (3.1), we have the following theorem.

THEOREM 3. Under (B), (G), (S) or (A), the likelihood ratio, Wald and score statistic are asymptotically equivalent. Under H_o,

$$\lambda_n, w_n \text{ and } c_n \underset{d}{\to} \chi^2(r) \ .$$

REMARK. For certain sequences $\{\beta(n)\}$ of alternatives, for which

$$\delta^2 = (C\beta(n) - \xi)'[CF_n^{-1}(\beta)C']^{-1}(C\beta(n) - \xi)$$

is constant, it is shown in Fahrmeir (1986) that

$$\lambda_n, w_n \text{ and } c_n \underset{d}{\to} \chi^2(r;\delta^2) \ .$$

4.1 BOUNDED REGRESSORS. The required set of conditions is:

(B)(i) *Bounded regressors*: $\|z_n\| < c$ for all n.

(ii) *Divergence*: $\lambda_{min} \sum_{i=1}^{n} z_i z_i' \to \infty$.

The divergence condition (ii), which is equivalent to $(\sum z_i z_i')^{-1} \to 0$, is the essential remaining assumption. In the classical linear regression model with i.i.d. errors, (ii) alone, without the boundedness assumption (i), is necessary and sufficient for weak (Drygas, 1976) and strong (Lai, Robbins and Wei, 1979) consistency.

For $z_n' = (1, x_n')$, in particular in the cumulative logit model, condition (B)(ii) is, under (B)(i), equivalent to

(B)(ii)' *Divergence of the scatter matrix*:

$$\lambda_{min} \sum_{i=1}^{n} (x_i - \bar{x}_n)(x_i - \bar{x}_n)' \to \infty, \qquad \bar{x}_n = \sum_{i=1}^{n} x_i/n \quad .$$

Thus one might expect satisfactory finite sample behaviour of the MLE and the test statistics, if the sequence $\{x_n\}$ of regressors is sufficiently scattered. Indeed, simulation results sustain this conjecture even for moderate sample sizes.

4.2 GROWING REGRESSORS. Although bounded regressors cover a large number of situations, there are applications where models with growing sequences $\{z_n\}$ of regressors will be of interest, e.g. to model certain trends in a longitudinal analysis. The following set of conditions presents, under a somewhat sharpened divergence condition, admissible growth rates which permit asymptotical inference. For convenience, the linear Poisson model is treated separately.

(G1)(i) *Admissible growth rates*:

$$\|z_n\| = o(\log n) \quad (\text{resp. } \|x_n\| = o(\log n))$$

for the binary, multinomial and cumulative logit models, and for the loglinear Poisson model,

$$\|z_n\|^2 = o(\log n)$$

for the probit model.

(ii) *Divergence of order* n^α: for some $\alpha > 0$, $c > 0$, $n_1 \in \mathbb{N}$

$$\lambda_{\min} \sum_{i=1}^n z_i z_i' \geq c\, n^\alpha \quad , \qquad n \geq n_1 \; .$$

For $z_n' = (1, x_n')$, in particular in the cumulative logit model, (G1)(ii) is, under (G1)(i), equivalent to the corresponding divergence of the scatter matrix:

(G1)(ii)' For some $\alpha > 0$, $c > 0$, $n_1 \in \mathbb{N}$

$$\lambda_{\min} \sum_{i=1}^n (x_i - \bar{x}_n)(x_i - \bar{x}_n)' \geq c\, n^\alpha ,\; n \geq n_1 \; .$$

For the linear Poisson model we require

(G2)(i) $\quad \|z_n\| = O(n^\alpha) \quad$ for some $\alpha > 0$,

(ii) $\quad \lambda_{\min} \sum_{i=1}^n z_i z_i' \geq c\, n^\alpha \; , \quad n \geq n_1$

for some $c > 0$, $n_1 \in \mathbb{N}$, with the same α as in (i).

In particular, the admissible growth rates given above imply that the Fisher information diverges, which seems to be an indispensable requirement for asymptotic inference. Moreover they represent critical upper bounds in that the Fisher matrix converges if they are exceeded monotonically:

REMARK. (i) For the binary logit model the following partial converse holds: For parameters β such that

$$|z_n'\beta| \geq c \log n \; , \qquad n \geq n_1$$

for some $c > 1$ and $n_1 \in \mathbb{N}$, $\lambda_{\min} F_n(\beta)$ converges. Similar results hold for the probit, multinomial and cumulative logit models and the loglinear Poisson model.

(ii) If $z_n' = (1, x_n')$ in the linear Poisson model, with $\|x_n\| > c\, n^\alpha$, $n \geq n_1$, for some $\alpha > 1$, $c > 0$, $n_1 \in \mathbb{N}$, then the divergence condition (G2)(ii) cannot hold. Furthermore, $\lambda_{\min} F_n(\beta)$ converges for all $\beta \in B$.

The results may be illustrated by a sequence $\{(1, x_n)\}$ of regressors, where x_n is scalar.

EXAMPLE. (i) Let $z_n' = (1,x_n)$ be given by $x_n = (\log n)^{\alpha}$. For the binary logit model, (G1) holds if $0 < \alpha < 1$. If $\alpha \geq 1$, then $F_n(\beta)$ converges for some β.

(ii) For the linear Poisson model let x_n in $z_n' = (1,x_n)$ be given by n^{α}. Then (G2) holds if $0 < \alpha \leq 1$, whereas $F_n(\beta)$ converges for any β if $\alpha > 1$.

Compared to classical linear regression, these results seem to be somewhat disappointing: It is well known that in classical linear regression higher growth rates, in particular polynomial trends, are admissible for asymptotical inference. In the models under consideration, only rather limited growth rates are admissible. How can we interpret this? For example, look at the binary logit model. The probabilities π_n will tend to one resp. zero, if some regressors are growing monotonically to $+\infty$ resp. $-\infty$. Thus for large n nearly all responses y_n will fall into one category and there will be to less information to draw inferences about the relevant parameters. Analogous arguments can be given for the other models. The admissible growth rates in (G1)(i) , (G2)(i) assure that enough information is available and that asymptotic theory works.

Simulation results substantiate these reasonings, see Section 5 for the binary logit model.

Finally let us compare the admissible growth rates of the binary logit and probit model: for the probit model the admissible growth rate is $\|z_n\| = o(\log n)^{1/2}$, i.e. even lower than for the logit model. This fact is to some extent in contrast to recommondations in the literature which state that there are essentially only computational differences between logit and probit models. It reflects the different asymptotic behaviour of the logistic and the normal density as $|t| \to \infty$: the logistic density behaves like $\exp(-|t|)$, the probit model like $\exp(-|t|^2)$.

4.3 STOCHASTIC REGRESSORS. Here we assume

(S)(i) The pairs $\{(y_n,z_n)\}$ resp. $\{(y_n,x_n)\}$ are i.i.d. as a pair (y,z) resp. (y,x) of random variables.

Under weak additional assumptions on the (marginal) distribution of z resp. x Theorems 1,2,3 remain valid:

(S)(ii) The matrix of second moments of z resp. the covariance matrix of x exists and is positive definite:

$$0 < E(zz') < \infty \quad , \quad \text{resp.}$$

$$0 < \text{cov}(x) < \infty \quad , \quad \text{if } z' = (1,x') \ .$$

(iii) Additionally,

a) $E\,\|z\|^4 < \infty$ for the probit model,

b) $E \exp z'\beta < \infty$ for all β, i.e. the moment generating function exists, for the loglinear Poisson model,

c) $B = \text{int}\{\beta : z'\beta > 0 \text{ a.s.}\}$ is nonvoid for the linear Poisson model, and the true parameter β lies within this B.

Again, the additional moment condition $E\,\|z\|^4 < \infty$ for the probit model reflects the fact, that the logistic density has a heavier tail than the normal density.

4.4 GENERAL CONDITIONS. In this subsection we give general conditions under which Theorems 1,2 and 3 hold. They can be used to deduce the results of the preceding subsections 4.1 to 4.3, as well as to treat situations which are not covered by these subsections, e.g. mixed deterministic/ stochastic regressors. The assumptions are:

(A)(i) *Divergence of the Fisher information*:

$$\lambda_{min}\, F_n(\beta) \to \infty \ .$$

(ii) *Negligibility conditions*:

a) $$z_n'\, F_n^{-1}(\beta) z_n \to 0$$

for the binary logit and the loglinear Poisson model,

b) $$\text{tr}\ Z_n' F_n^{-1}(\beta) Z_n \to 0$$

for the multinomial and cumulative logit model,

c) $\|z_n\|^2 \, z_n' \, F_n^{-1}(\beta) z_n \to 0$

for the probit model.

For the linear Poisson model no negligibility condition is necessary, and (A)(i) is equivalent to

$$\text{(A)(i)}' \qquad \lambda_{\min} \textstyle\sum_{i=1}^{n} z_i z_i' / \|z_i\| \to \infty \, .$$

REMARKS. (i) Conditions (A)(i) and (A)(ii)a) are equivalent to

$$\max_{1 \le i \le n} z_i' \, F_n^{-1}(\beta) z_i \to 0 \qquad (4.2)$$

alone. This holds analogously for b) and c). Thus the negligibility conditions have the same interpretation as the Feller condition in the Lindeberg-Feller central limit theorem: Regressors must not be too large compared to the information cumulated in $F_n(\beta)$.

(ii) Condition c) for the probit model is clearly stronger than a), reflecting again the different behaviour of the tails of logistic and normal densities. This corrects the statement of F/K (in the introduction and at the end of p.362) that (4.2) is a sufficient condition for consistency and asymptotic normality of the MLE in the probit model too. The arguments on p.362 leading to this statement do not apply, since u is not bounded, and a) has to be replaced by the somewhat stronger condition c).

(iii) Conditions (A) seem to be near to necessity for the assertions of Theorems 1-3. In particular, divergence of the Fisher information seems to be indispensable. A simple example is provided by the linear Poisson model with scalar regressors $z_n > 0$, $n \in \mathbb{N}$. Then $B = \{\beta > 0\}$, and the MLE can be given explicitly by $\hat{\beta}_n = \sum_1^n y_i / v_n$, where $v_n = \sum_1^n z_i$. Condition (A)(i)' reduces to $v_n \to \infty$. Since $v_n \hat{\beta}_n$ is Poisson distributed with rate $v_n \beta$, it is easy to see that $v_n \to \infty$ is also necessary for $P(\hat{\beta}_n \in B) \to 1$, consistency and asymptotic normality of $\{\hat{\beta}_n\}$.

5. SIMULATION RESULTS FOR FINITE SAMPLES. To get some insight into the relevance of asymptotic theory for finite samples, Monte Carlo simulations have been carried out. We present results for the binary logit model in the critical case of a monotonically growing sequence of scalar regressors

$$z_i = (\log i)^{\alpha}\ , \quad i=1,\ldots,n\ ,$$

for various values of α, $0 < \alpha \leq 2$, and values of 1,2 and 3 for the parameter β. Recall that, due to Subsection 4.3 the growth rate is admissible for $\alpha < 1$, whereas for $\alpha \geq 1$ the Fisher information converges. For each combination of α and β, 100 simulation runs of sample size $n = 200$ have been carried out. For every run the MLE $\hat{\beta}_n$ was computed if it existed, and - for fixed (α,β) - the average $\bar{\beta}$ and empirical variance s^2 of existing MLE's as well as the asymptotic variance σ^2 obtained from the Fisher information and the estimated mean square error (MSE) were evaluated.[1)]

Figure 1 displays the relative frequencies $f_n = f_n(\alpha,\beta)$ of existence of the MLE. Asymptotically, we have $f_n \to 1$ for $\alpha < 1$ and $f_n \to f < 1$ for $\alpha \geq 1$. For finite n there is no clear cut off point at $\alpha = 1$, but the tendency of $f_n(\alpha,\beta)$ getting small for $\alpha > 1$ is clearly exhibited.

Figure 2 compares $\bar{\beta} = \bar{\beta}(\alpha,\beta)$ with the corresponding true values $\beta = 1,2$ and 3. For $\alpha > 1$ bias increases drastically, in particular for $\beta = 3$.

Figure 3 shows, for $\beta = 2$ only, $s^2, \hat{\sigma}^2$ and MSE. For $\alpha < 1$, all three curves should coincide asymptotically. For finite sample size, there is a strongly increasing difference for growing α.

These results clearly indicate that statistical analysis will become unreliable if critical growth rates are reached or exceeded monotonically. However, if the regressors are sufficiently scattered, e.g. for alternating sequences or for stochastic regressors, finite sample behaviour is much better.

---

1) We thank Wolfgang Hennevogl for the computations. They were carried out on an Olivetti PC M24, using a first version of GLAMOUR, a program package developed at the University of Regensburg, supported in part by the DFG.

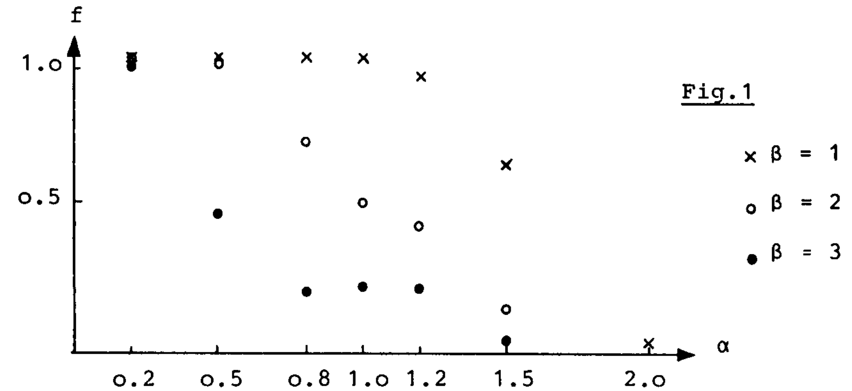

Fig.1

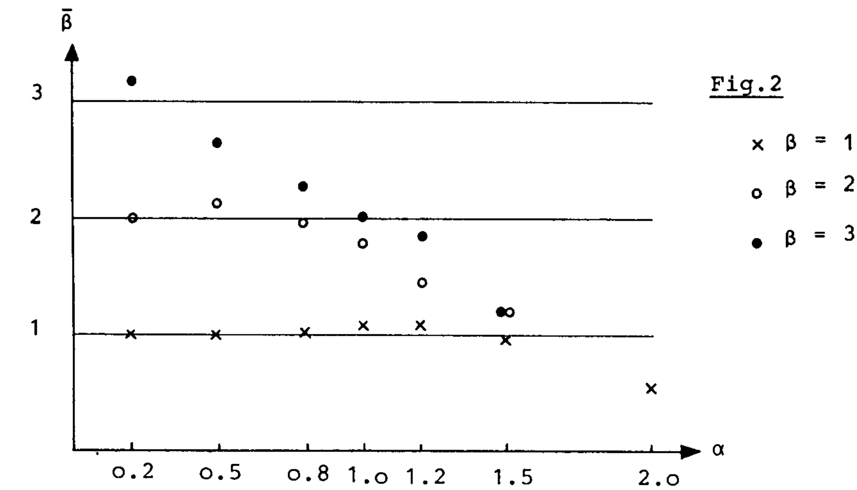

Fig.2

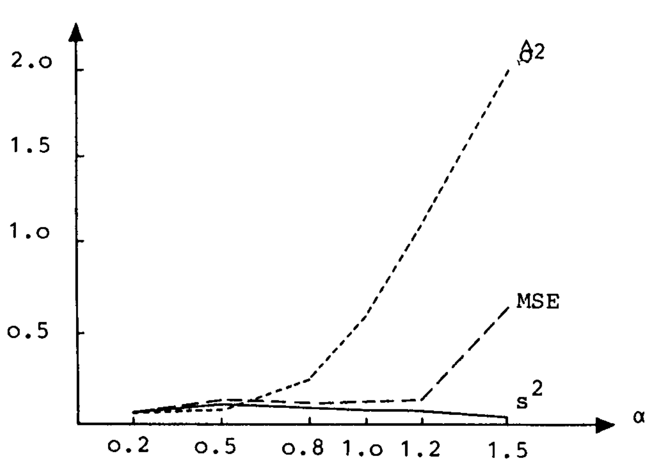

Fig.3

6. PROOFS. For the purposes of this section and for easier reference to results in F/K it is more convenient to formulate the multinomial (binomial for q=1) models in terms of the natural parameter $\theta_n = (\theta_{n1},\ldots,\theta_{nq})'$ instead of $\pi_n = (\pi_{n1},\ldots,\pi_{nq})'$. The relation between θ_n and π_n is given by (see e.g. Fahrmeir, Hamerle 1984,p.45) $\theta_{nj} = \log(\pi_{nj}/\pi_{nm})$. The model equation (2.9) for multinomial responses, i.e. $\pi_n = h(Z_n'\beta)$, transforms into $\theta_n = u(Z_n'\beta)$, where $u(\gamma) = (u_1(\gamma),\ldots,u_q(\gamma))'$ is obtained by insertion of $\pi_{nj} = h_j(\gamma)$, $j=1,\ldots,q$, into (6.1).

For the binomial and multinomial logit model, the function u is simply the identity, whereas

$$u(\gamma) = \log\phi(\gamma) - \log(1-\phi(\gamma)) \tag{6.1}$$

for the probit model, and

$$u_j(\gamma_1,\ldots,\gamma_q) = \log(1+\exp(\gamma_q)) + \log[(1+\exp(-\gamma_j))^{-1} - (1+\exp(-\gamma_{j-1}))^{-1}] , \tag{6.2}$$

$j=1,\ldots,q$, $\gamma_o = -\infty$, for the cumulative logit model. The log likelihood and its derivatives, the Fisher matrix, etc., can be expressed in terms of u and its derivatives $\dot{u}(\gamma) = \partial u(\gamma)/\partial\gamma, \ddot{u}(\gamma) = \partial^2 u(\gamma)/\partial\gamma\partial\gamma'$, as in F/K p.347; e.g.

$$F_n(\beta) = \textstyle\sum_{i=1}^n Z_i U_i(\beta)\Sigma_i(\beta)U_i'(\beta)Z_i' ,$$

$$H_n(\beta) = F_n(\beta) - \textstyle\sum_{i=1}^n\sum_{j=1}^q Z_i W_{ij}(\beta)Z_i'(y_{ij}-\pi_{ij}(\beta)) ,$$

with

$$U_i'(\beta) = \dot{u}(Z_i'\beta),\ W_{ij}(\beta) = \ddot{u}_j(Z_i'\beta),\ j=1,\ldots,q. \tag{6.3}$$

In Subsections 4.2 - 4.4, the following properties will be utilized: For the probit model

$$\dot{u}(\gamma) = \varphi(\gamma)/\phi(\gamma)(1-\phi(\gamma)) ,$$

where φ is the standard normal density. Using an asymptotic expansion of $1-\phi(\gamma)$ (Abramowitz and Stegun, 1965, p.932, formula 26.2.12), it can be seen that for $|\gamma| \to \infty$ $\dot{u}(\gamma)$ behaves like $|\gamma|$ and

$$0 < c \le |\dot{u}(\gamma)| \le a + |\gamma| \tag{6.4}$$

for some $a > 0$, $c > 0$. Furthermore, second derivatives are bounded and uniformly continuous.

For the cumulative logit model it can be verified by some lengthy but straightforward calculations that $U_i(\beta)$, $W_{ij}(\beta)$ are bounded and continuous, uniformly in i, and that $\lambda_{\min} U_i(\beta)U_i'(\beta)$ is bounded away from zero.

6.1 PROOFS TO SUBSECTION 4.1. Sufficiency of condition (B) for condition (A) can be verified by the arguments of F/K for regressors with a compact range. First one shows that, under (B)(i), (B)(ii) implies (A)(i). The negligibility conditions (A)(ii) follow immediately. The equivalence of (B)(ii) and (B)(ii)' follows, under (B)(i), by an application of the matrix inversion lemma.

6.2. PROOFS TO SUBSECTION 4.2. CONDITIONS (G) IMPLY (A). *Binary logit model.* The information matrix is given by $F_n(\beta) = \sum_1^n z_i z_i' \sigma_i^2$, with

$$\sigma_n^2 = \pi_n(1-\pi_n) = (1+e^{-z_n'\beta})^{-1}(1+e^{z_n'\beta})^{-1} , \qquad (6.5)$$

according to the model equation (2.1). For the sufficiency part as well as for the partial converse given in Remark (i), the inequality

$$e^{-|\gamma|}/4 \le (1+e^{-\gamma})^{-1}(1+e^{\gamma})^{-1} \le e^{-|\gamma|} \qquad (6.6)$$

will be of advantage. The admissible growth rate $\|z_n\| = o(\log n)$ in (G1)(i) is equivalent to

$$\exp(-\|z_n\|\,\|\beta\|) \ge n^{-\delta} \quad \text{for all } \beta,\ \delta > 0,\ n \ge n_1(\beta,\delta).$$

In combination with (6.5) and the left inequality in (6.6), $\lambda_{\min} F_n(\beta)$ can be estimated from below:

$$\lambda_{\min} F_n(\beta) \ge c_1 + \frac{1}{4}\lambda_{\min} \sum_1^n z_i z_i' i^{-\delta}$$

with some constant $c_1 = c_1(\delta)$. Since $i^{-\delta} \ge n^{-\delta}$ for $i \le n$, $\delta > 0$, under (G1)(ii) we have

$$\lambda_{\min} F_n(\beta) \ge \frac{c}{4} n^{\alpha-\delta} + c_1 , \qquad n \ge n_1 .$$

With $\delta = \alpha/2$ this implies (A)(i); more precisely, $\lambda_{\min} F_n(\beta)$ diverges at least as $n^{\alpha/2}$. Since $\|z_n\|^2 = o[(\log n)^2]$,

$$z_n' F_n^{-1} z_n \le \|z_n\|^2/\lambda_{min} F_n(\beta) \to 0 \ .$$

Thus the negligibity condition (A)(ii) holds.

Loglinear Poisson model. It is considered as an example in F/K for a scalar sequence $\{z_n\}$. The extension to higher dimensions of $\{z_n\}$ resembles strongly the proof for the binary logit model and is omitted.

Probit model. Now $F_n(\beta) = \sum_1^n z_i z_i' (\dot{u}_i)^2 \sigma_i^2$, with $\sigma_i^2 = \phi(z_i'\beta)(1-\phi(z_i'\beta))$. Since $\dot{u}$ is bounded away from zero and

$$\phi(\gamma)(1-\phi(\gamma)) \ge c \exp(-\gamma^2)$$

for some $c > 0$, the same reasoning as for the logit model applies with $\|z_n\|^2$ instead of $\|z_n\|$.

Multinomial logit model. We have $F_n(\beta) = \sum_1^n Z_i \Sigma_i Z_i'$, $\Sigma_n = \operatorname{diag}(\pi_n) - \pi_n \pi_n', \pi_n = (\pi_{n1}, \dots, \pi_{nq})'$. Gershgorin's Theorem (e.g. Stoer, Bulirsch, 1973, p.73) allows to bound $\lambda_{min}\Sigma_n$ from below:

$$\lambda_{min}\Sigma_n \ge \min(\pi_{n1}\pi_{nm}, \dots, \pi_{nq}\pi_{nm}) \ .$$

Inserting the model equations (2.3), it can be seen that

$$\lambda_{min}\Sigma_n \ge c \exp(-3(|\gamma_{n1}|+\dots+|\gamma_{nq}|), \ \gamma_{nj} = z_n'\beta_j \ ,$$

$c > 0$ some appropriate constant. With the same arguments as for the binary logit model, one arrives at

$$\lambda_{min} F_n(\beta) \ge k n^{\alpha-\delta} + c_1, \ n \ge n_1 \ ,$$

Z_i as in (2.10). The rest of the proof follows by noting that $\lambda_{min} \sum Z_i Z_i' = \lambda_{min} \sum z_i z_i'$ and $\operatorname{tr} Z_n' F_n^{-1} Z_n \le q\|z_n\|^2/\lambda_{min} F_n$.

Cumulative logit model. This can be treated by analogous arguments; note that $\lambda_{min} U_i(\beta) U_i'(\beta)$ is bounded away from zero and $\lambda_{min} \sum Z_i Z_i' \ge q \, \lambda_{min} \sum z_i z_i'$, $z_i' = (1, x_i')$.

Linear Poisson model. Here the assertion can be shown with the following analogue of the Kronecker lemma.

LEMMA 1. Let $\{a_n\}$ be a sequence with $a_n - a_{n-1} \ge 0$, $a_n > 0$ and $a_n \to \infty$. Then $\lim \lambda_{min} \sum_{i=1}^n a_i^{-1} z_i z_i' < \infty$ implies that $a_n^{-1} \lambda_{min} \sum_1^n z_i z_i' \to 0$.

PROOF. Compare the proof of the Kronecker lemma given e. g. in Chung (1974). Letting $B_n = \sum_1^n a_i^{-1} z_i z_i'$, $b_n = \lambda_{min} B_n$, it holds that

$$a_n^{-1} \sum_1^n z_i z_i' = B_n - a_n^{-1} \sum_1^{n-1} B_i(a_{i+1}-a_i) \; ,$$

$$0 \le a_n^{-1} \lambda_{min} \sum_1^n z_i z_i' \le b_n - a_n^{-1} \sum_1^{n-1} b_i(a_{i+1}-a_i) \; .$$

The rightmost expression converges to zero, whence the assertion of Lemma 1.

The growth condition $\|z_n\| = O(n^{\alpha})$ implies

$$\lambda_{min} \sum_1^n \|z_i\|^{-1} z_i z_i' \ge c_1 \lambda_{min} \sum_1^n i^{-\alpha} z_i z_i' \; ,$$

with some constant $c_1 > 0$. With $a_n = n^{\alpha}$, using that (G2)(ii) requires $a_n^{-1} \lambda_{min} \sum_1^n z_i z_i'$ asymptotically to be bounded away from zero, Lemma 1 implies that the right side of the inequality above diverges. Divergence of the left side is equivalent to condition (A), for the linear Poisson model.

PROOF OF THE REMARK. *Binary logit model.* The right side of (6.6) allows to bound $\lambda_{min} F_n(\beta)$ from above:

$$\lambda_{min} F_n(\beta) \le c_o + \sum_{n_1}^n |z_i'\beta|^2 e^{-z_i'\beta} \le c_o + c^2 \sum_{n_1}^n (\log i)^2 i^{-c}$$

for sufficiently large n_1 and some $c_o > 0$. For $c > 1$ the right side of the inequality, and thus $\lambda_{min} F_n(\beta)$, converges.

Linear Poisson model. If $z_n' = (1, x_n')$, since $\lambda_{min} A \le \min(a_{11}, \ldots, a_{pp})$ for any symmetric p×p-matrix $A = (a_{ij})$, we have $\lambda_{min} \sum z_i z_i' \le n$, whence (G2)(ii) cannot hold with some $\alpha > 1$. With the same argument,

$$\lambda_{min} \sum \|z_i\|^{-1} z_i z_i' \le \sum_1^n \|z_i\|^{-1} \; .$$

If $\|x_n\| \ge c\, n^{\alpha}$, for all $n > n_1$, with some $c > 0$, $\alpha > 1$, $n_1 \in \mathbb{N}$, the series $\sum_1^n \|z_i\|^{-1}$ converges.

PROOFS TO THE EXAMPLE. *Binary logit model, sufficiency part:* if $z_n = (1, x_n)$ with $x_n = (\log n)^{\alpha}$, $0 < \alpha < 1$, clearly $\|z_n\| = o(\log n)$. It remains to show that (G1)(ii)' holds. L'Hopital's rule can be used to demonstrate, for any real α,

$$\int_2^n (\log x)^{\alpha}\, dx / (n(\log n)^{\alpha}) \to 1 \; . \tag{6.7}$$

With some c_1 depending only on α, we have

$$\int_2^{n-1} (\log x)^\alpha \, dx \le \sum_2^{n-1} (\log i)^\alpha \le c_1 + \int_2^n (\log x)^\alpha \, dx, \tag{6.8}$$

for any real α. Using (6.7), (6.8) and the formula

$$\int_1^n (\log x)^\alpha \, dx = n(\log n)^\alpha - \int_1^n (\log x)^{\alpha-1} \, dx, \quad \alpha > 0 ,$$

it follows that

$$(\log n)^{1-\alpha}(x_n - \bar{x}_{n-1}) \to \alpha, \qquad \alpha > 0 . \tag{6.9}$$

From (6.9) and

$$\sum_{i=1}^n (x_i - \bar{x}_n)^2 = \sum_{i=2}^n \frac{i-1}{i}(x_i - \bar{x}_{i-1})^2 ,$$

utilizing the theorem of Stolz (e.g. Loève, 1977), we obtain

$$\frac{\sum_1^n (x_i - \bar{x}_n)^2}{\sum_2^n (\log i)^{2\alpha-2}} \to \alpha^{-2} .$$

Using (6.8), (6.7) again, we can replace the denominator by $n(\log n)^{2\alpha-2}$. This implies (G1)(ii)'.

The converse part for the binary logit model is immediate from the remark. The assertions for the linear Poisson model can be shown with similar considerations, whence the proof is omitted.

6.3 PROOFS TO SUBSECTION 4.3. For stochastic regressors it suffices to show that condition (A) holds almost surely under (S). Since arguments are similar to those of F/K, proofs are only indicated. Assumptions (S)(ii),(iii) assure that the expected information E F(β) exists and is positive definite, where F(β) denotes the information of a single observation y conditional on Z.

For instance, $F(\beta) = zz' \exp(z'\beta)$ for the loglinear Poisson model, and finiteness of E F(β) is equivalent to $E\|z\|^2 \exp(z'\beta) < \infty$. The assumption $E \exp z'\beta < \infty$, for all β, is equivalent to $E \exp(\|z\|\lambda) < \infty$, for all λ. This can be seen by choosing, for a given λ, a finite number $\beta_1, \dots, \beta_m$ of vectors such that $\max_{i=1,\dots,m} z'\beta_i \ge \|z\|\lambda$, for any vector z (such a choice is always possible). Then $E \sum_1^m \exp(z'\beta_i) < \infty$ implies $E \exp(\|z\|\lambda) < \infty$ (the other direction follows by the Cauchy-Schwarz inequality).- For any given β, $\exp(\|z\|\lambda)$ dominates $\|z\|^2 \exp(z'\beta)$, with an appropriate λ. Hence $E\,F(\beta)$ is finite, if the moment generating function of z exists. Under finiteness, $E\,F(\beta)$, as well as Ezz', is positive definite if and only if $P(z'\lambda \neq 0) > 0$, for any $\lambda \neq 0$. For the other models, similar arguments apply.

By Kolmogorov's strong law for i.i.d. sequences,

$$F_n(\beta)/n \overset{a.s.}{\to} E\,F(\beta) \ ,$$

implying (A)(i), with $\lambda_{min} F_n(\beta)$ growing as fast as n. A further application of the strong law then shows that (A)(ii) is fulfilled: E.g. for (A)(ii)a)

$$\sum_{i=1}^n z_i z_i'/n \overset{a.s.}{\to} Ezz' \ ,$$

hence $\|z_n\|^2/n \to 0$, and (A)(ii)a) follows.

6.4 PROOFS TO SUBSECTION 4.4. In this section β_o denotes the true but unknown parameter which is to be estimated, and β is any parameter in B. For notational simplicity we shall mostly drop the argument β_o in $s_n(\beta_o)$, $F_n(\beta_o)$, P_{β_o}, E_{β_o} etc. and write s_n, F_n, E, P etc.

To prove that the statements of Theorems 1,2 and 3 hold under (A), we will show that (A)(i) and (A)(ii) imply asymptotic normality of the score function,

$$F_n^{-1/2} s_n \underset{d}{\to} N(0,I) \tag{6.10}$$

and, in the same notation as in F/K, the following conditions:

(N*) for all $\delta > 0$

$$\max_{\beta \in N_n(\delta)} \|F_n^{-1/2} H_n(\beta) F_n^{-T/2} - I\| \underset{p}{\to} 0 \ ,$$

where $N_n(\delta) = \{\beta : \|F_n^{T/2}(\beta-\beta_o)\| \leq \delta\}$, n = 1,2,... is a sequence of neighborhoods of β_o.

(N) $\max_{\beta \in N_n(\delta)} \|F_n^{-1/2} F_n(\beta) F_n^{-T/2} - I\| \to 0$ for all $\delta > 0$.

Note that (N*) reduces to (N) for models with $F_n(\beta) = H_n(\beta)$, i.e. the binary and multinomial logit model, and the loglinear Poisson model. Together with the divergence condition (A)(i), the assumptions (6.10) (N*) and (N) imply that Theorems 1,2 and 3 hold: Theorem 1 follows directly from F/K (p.349,p.360). Additionally the LAN (local asymptotic normality) condition holds, see Lemma 1 of Fahrmeir (1986). By the same arguments as in Kaufmann (1987, Theorem 2), Theorem 2 follows. Theorem 3 and the remark are an immediate consequence of Fahrmeir (1986).

It remains to show that (A) implies (6.10), (N*) and (N). The binary and multinomial logit model, and the loglinear Poisson model have been considered in Section 3 of F/K as well as by Haberman (1977a,b). To give a unified proof for the probit model and the cumulative logit model, we first state a lemma which adresses to multinomial response models in general. The functions U_n, W_{nj} appearing in the conditions (M) below are defined in (6.3).

LEMMA 2. Suppose that the following conditions (M) hold for a multinomial response model.

(M)(i) $\lambda_{min} F_n \to \infty$,

(ii) $\operatorname{tr} U_n' Z_n' F_n^{-1} Z_n U_n \to 0$,

(iii) $\operatorname{tr} \Sigma_n^{-1} \operatorname{tr} \Sigma_n$ is bounded, uniformly in n,

(iv) The first and second derivatives $\dot{u}$, $\ddot{u}$ exist, $\lambda_{min} U_n U_n'$ is bounded away from zero, $W_{nj}(\beta)$, $j=1,\dots,q$, is bounded and continuous, uniformly in n.

Then the normed score function is asymptotically normal, i.e. (6.10) holds, and conditions (N*) and (N) are fulfilled.

COROLLARY 1. For the cumulative logit model, condition (A)(ii)b), i.e. $\operatorname{tr} Z_n' F_n^{-1} Z_n \to 0$, and for the probit model, condition (A)(ii)c), i.e. $\|z_n\|^2 z_n' F_n^{-1} z_n \to 0$, imply M(ii), (iii),(iv). Thus, Lemma 2 holds if additionally the divergence condition M(i) is fulfilled.

PROOF OF LEMMA 2. (i) Asymptotic normality of the score function can be shown, with minor modifications, as in the proof of Lemma 2 of F/K, p.363, using the Lindeberg-Feller theorem.

(ii) Conditions (N*) and (N) can be verified via the same decompostion of $F_n^{-1/2}H_n(\beta)F_n^{-T/2}-I$ as in the proof of Theorem 4 of F/K. We pick out the terms

$$A_n(\beta) = F_n^{-1/2}F_n(\beta)F_n^{-T/2} - I =$$

$$= \sum_{i=1}^n F_n^{-1/2}Z_i(U_i(\beta)\Sigma_i(\beta)U_i'(\beta) - U_i\Sigma_iU_i')Z_i'F_n^{-T/2}$$

and

$$C_n(\beta) = \sum_{i=1}^n \sum_{j=1}^q F_n^{-1/2}Z_i(W_{ij}(\beta) - W_{ij})Z_i'F_n^{-T/2}(y_{ir} - \pi_{ij}) \ .$$

First note that

$$\sum_1^n \lambda' F_n^{-1/2}Z_iU_i'\Sigma_iU_i'Z_i'F_n^{-T/2}\lambda = 1 \tag{6.11}$$

for any vector λ with $\lambda'\lambda = 1$. For any such λ we have, with

$$M_i(\beta) = \Sigma_i^{-1/2}U_i^{-1}U_i(\beta)\Sigma_i(\beta)U_i'(\beta)U_i^{-T}\Sigma_i^{-T/2} - I \ ,$$

$$\lambda' A_n(\beta)\lambda = \sum_1^n \lambda F_n^{-1/2}Z_iU_i\Sigma_i^{1/2}M_i(\beta)\Sigma_i^{T/2}U_i'F_n^{-T/2}\lambda$$

and, in view of (6.11),

$$\max_{1\le i\le n,\beta\in N_n(\delta)} \lambda' A_n(\beta)\lambda \le \max_{i,\beta}\|M_i(\beta)\| \ . \tag{6.12}$$

Taylor expansion of $U_i(\beta)$ at β_o gives, for some $\kappa > 0$,

$$\|U_i^{-1}U_i(\beta) - I\| \le \kappa\|Z_i'(\beta-\beta_o)\| \le \kappa\|Z_i'F_n^{-T/2}\|\|F_n^{T/2}(\beta-\beta_o)\| \ ,$$

since $W_{ij}(\beta)$ is bounded and $\lambda_{min}U_iU_i'$ is bounded away from zero. By definition of $N_n(\delta)$

$$\max_{i,\beta}\|U_i^{-1}U_i(\beta)-I\| \le \kappa\ \delta \max_i \|Z_i'F_n^{-T/2}\| \to 0, \tag{6.13}$$

since $\max_i\|Z_i'F_n^{-T/2}\| \to 0$ from M(i), M(ii) and M(iv). Similarly,

$$\max_{i,\beta}\|\Sigma_i^{-1/2}\ \Sigma_i(\beta)\ \Sigma_i^{-T/2} - I\| \to 0 \tag{6.14}$$

can be seen, using the last inequality in the proof of

Lemma 5 on p.831 of Haberman (1977a). Now, with (6.13), (6.14) and M(iii), it follows easily that the right side and hence the left side of (6.12) converge to zero. This gives the result for $A_n(\beta)$:

$$\max_{\beta\in N_n(\delta)} \|A_n(\beta)\| \to 0 ,$$

and (N) is verified.

For $\max_{\beta\in N_n(\delta)}\|C_n(\beta)\|$, convergence to zero can be demonstrated in the first mean:

$$E \max_{\beta\in N_n(\delta)} |\lambda' C_n(\beta)\lambda| \le [\sum_{i=1}^{n} \|\lambda' F_n^{-1/2} Z_i U_i \Sigma_i^{1/2}|^2] \cdot \|U_i^{-1}\|^2$$

$$\cdot \max_{i,r,\beta}\|W_{ir}(\beta) - W_{ir}\| \cdot \max_i\{\|\Sigma_i^{-1}\| E \sum_{r=1}^{q} |y_{ir} - \pi_{ir}|\} \quad .$$

The first term of the product on the right side is 1, see (6.11); the second term is bounded from above, since $\lambda_{min} U_i U_i'$ is bounded away from zero by assumption (M)(iv); the third term tends to zero, due to (M)(i),(ii) and (iv). It remains to show that the last term is bounded: From $E|y_{ir} - \pi_{ir}| = 2\pi_{ir}(1-\pi_{ir})$ we get

$$E \sum_{r=1}^{q} |y_{ir} - \pi_{ir}| = 2 \text{ tr } \Sigma_i \quad .$$

Together with $\|\Sigma_i^{-1}\| = \lambda_{max}\Sigma_i^{-1} \le \text{tr } \Sigma_i^{-1}$, the conclusion follows from assumption (M)(iii). Treating B_n and $D_n(\beta)$ similarly, (N*) is verified.

PROOF OF COROLLARY 1. *Probit model.* Condition (M)(iii) is valid for any binary model, since $\Sigma_n = \sigma_n^2$ is scalar, and $\dot{u},\ddot{u}$ have the properties required in (M)(iv), see the beginning of this section. Finally,(M)(ii) follows from (6.4) and (A)(ii)c).

Cumulative logit model. Condition M(iv) follows directly from the properties of U_n, W_{nj} stated at the beginning of Section 6, and (A)(ii)b) implies (M)(ii), since U_n is bounded. It remains to check (M)(iii). From $\Sigma_n = \text{diag}(\pi_n) - \pi_n\pi_n'$, we have

$$\mathrm{tr}\,\Sigma_n = \pi_{n1}(1-\pi_{n1})+\ldots+\pi_{nq}(1-\pi_{nq}) ,$$

$$\Sigma_n^{-1} = \mathrm{diag}\,(\frac{1}{\pi_{n1}},\ldots,\frac{1}{\pi_{nq}}) + \frac{1}{\pi_{nm}}\,\mathbb{1}\,\mathbb{1}', \quad \mathbb{1}' = (1,\ldots,1),$$

$$\mathrm{tr}\,\Sigma_n^{-1} = \frac{1}{\pi_{n1}} +\ldots+ \frac{1}{\pi_{nq}} + \frac{q}{\pi_{nm}} .$$

Inserting the model equations (2.4) for $\pi_{n1},\ldots,\pi_{nq}$, it is seen after some manipulations that the product $\mathrm{tr}\,\Sigma_n^{-1}\,\mathrm{tr}\,\Sigma_n$ is bounded.

Finally, we verify (A)(i), (6.10), (N*) and (N) for the *linear Poisson model*. We first discuss the assumption that B defined by (2.8) is nonvoid. More generally, we have the following lemma for an arbitrary index set I.

LEMMA 3. For p-vectors $z_i \neq 0$, $i \in I$, the set $\mathrm{int}\{\beta, z_i'\beta > 0, i \in I\}$ consists of all vectors β for which $\inf_{i\in I} z_i'\beta/\|z_i\| > 0$.

This lemma can be proved with similar considerations as in F/K, Subsection 3.4, proofs to Example (ii). It can be used to check whether B is nonvoid. For instance, if $z_n' = (1,x_n')$, then $\sup_n\|x_n\| < \infty$ (condition (B)(i)) is sufficient for B to be nonvoid, since at least $(1,0\ldots0) \in B$. If $z_n' = (1,x_n)$ with a scalar x_n, then B is nonvoid if and only if $\{x_n\}$ is bounded above or bounded below.

Lemma 3 provides the lower bound in

$$\gamma_1\|z_n\| \leq z_n'\beta \leq \gamma_2\|z_n\| , \qquad n \in \mathbb{N} , \tag{6.15}$$

which holds for all β in a sufficiently small neighbourhood of β_o, with constants γ_1,γ_2 depending only on β_o. This can be used to show that (A)(i) and (A)(i)' are equivalent, and to infer (6.10), (N*) and (N) from the single condition (A)(i)': With $\lambda_i = z_i'\beta_o$, we have

$$F_n = \sum \lambda_i^{-1} z_i z_i', \quad H_n = \sum y_i\lambda_i^{-2} z_i z_i' .$$

Clearly, (6.15) for $\beta = \beta_o$ implies that (A)(i)' and $\lambda_{min}\,F_n \to \infty$ are equivalent.

Further, for any unit vector λ, consider

$$\lambda' F_n^{-1/2} s_n = \textstyle\sum_{i=1}^n \alpha_{ni}(y_i-\lambda_i) ,$$

where $\alpha_{ni} = \lambda' F_n^{-1/2} z_i/\lambda_i$. Due to (6.15) and (A)(i)',

$$\max_{i=1,\dots,n} |\alpha_{ni}| \to 0 . \tag{6.16}$$

Together with $\sum \alpha_{ni}^2 \lambda_i = 1$ and the fact that y_i is Poisson distributed with rate λ_i, (6.16) suffices for $\sum_{i=1}^n \alpha_{ni}(y_i - \lambda_i) \to N(0,1)$. This can be shown making direct use of the characteristic or the moment generating function; the Lindeberg condition is too strong here.

Next, we have

$$v_n = \lambda' F_n^{-1/2} R_n F_n^{-T/2} \lambda = - \sum_{i=1}^n \alpha_{ni}^2 (y_i - \lambda_i), \quad Ev_n = 0 ,$$

$$\operatorname{var} v_n = \sum \alpha_{ni}^4 \lambda_i \leq \sum \alpha_{ni}^2 \lambda_i \max \alpha_{ni}^2 \to 0 .$$

Since λ is an arbitrary unit vector, this implies

$$F_n^{-1/2} H_n F_n^{-T/2} \to I \tag{6.17}$$

in quadratic mean. Using (6.17) and continuity arguments similar as in F/K, Subsection 3.4, proofs to example (ii), (N) and (N*) follow.

7. REFERENCES

Abramowitz,M. and Stegun,I.A. (1965). Handbook of Mathematical Functions. Dover, New York.

Albert,A. and Anderson,J.A. (1984). On the existence of maximum likelihood estimation in logistic regression models, Biometrika 71, 1-10.

Basawa,J.V. and Scott,D.J. (1983). Asymptotic Optimal Inference for Non-ergodic Models. Springer, Lecture Notes in Statistics, New York.

Chung,K.L. (1974). A course in probability theory. Academic Press, New York.

Drygas,H. (1976). Weak and strong consistency of the least squares estimators in regression models. Z. Wahrsch. verw. Gebiete 34, 119-127.

Fahrmeir,L. and Hamerle,A. (ed.)(1984). Multivariate statistische Verfahren. De Gruyter, Berlin.

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

Fahrmeir,L. (1986). Asymptotic testing theory for generalized linear models. Submitted to Math. Operationsforsch. Statist. Ser. Statist.

Gourieroux, C. and Monfort,A. (1981). Asymptotic properties of the maximum likelihood estimator in dichotomous logit models. J. Econometrics 17, 83-97.

Haberman,S.J. (1977a). Maximum likelihood estimates in exponential response models. Ann. Statist. 5, 815-841.

Haberman,S.J. (1977b). Log-linear models and frequency tables with small expected cell counts. Ann. Statist. 5, 1148-1169.
Haberman,S.J. (1980). Discussion of McCullagh's paper "Regression models for ordinal data", J.R. Statist. Soc. B42, 136-137.
Ibragimov,I.A. and Has'minskii,R.Z. (1981). Statistical Estimation, Asymptotic Theory. Springer, Berlin, Heidelberg, New York.
Jeganathan,P. (1982). On the asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal. Sankhyâ 44, Series A, 173-212.
Kaufmann,H. (1987). Regression models for nonstationary categorical time series: asymptotic estimation theory. Ann. Statist., to appear.
Kaufmann,H. (1986b). On directions of strictness, affinity and constancy of proper convex functions. Preprint.
Kaufmann,H. (1986c). On the uniqueness of the maximum likelihood estimator in quantal and ordinal response models. Submitted to Metrika.
Lai,T.L., Robbins,H. and Wei,C.Z. (1979). Strong consistency of least squares estimates in multiple regression II. J. Multivariate Anal. 9, 343-361.
Loève,M. (1977). Probability theory (4th ed.) Springer, Berlin, New York.
Mathieu,J.R. (1981). Tests of χ^2 in the generalized linear model. Math. Operationsforsch. Statist. Ser. Statist. 12, 509-527.
McCullagh,P. and Nelder,J.A. (1983). Generalized linear models. Chapman and Hall, London.
Nordberg,L.(1980). Asymptotic normality of maximum likelihood estimators based on independent, unequally distributed observations in exponential family models. Scand. J. Statist. 7, 27-32.
Silvapulle,M.J. (1981). On the existence of maximum likelihood estimators for the binomial response models. J.R. Statist. Soc. B43, 310-313.
Wedderburn,R.W.M. (1976). On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models, Biometrika 63, 27-32.

Prof. Dr. Ludwig Fahrmeir
Dr. Heinz Kaufmann

Universität Regensburg
Postfach
D-84oo Regensburg