The Annals of Statistics
1989, Vol. 17, No. 3, 1198–1216

BOOTSTRAPPING THE MAXIMUM LIKELIHOOD ESTIMATOR IN HIGH-DIMENSIONAL LOG-LINEAR MODELS[1]

BY WILHELM SAUERMANN

Gödecke AG

The notion of a bootstrap estimator of the distribution of the maximum likelihood estimator in log-linear models is defined for common sampling models. It is shown that the bootstrap estimator is consistent under assumptions which allow the dimension of the model to increase to infinity. Such an approach allows treatment of large, sparse contingency tables.

1. Introduction. The bootstrap [Efron (1979) and Beran (1984)] is a general, easily implemented method for obtaining estimates of the distribution of a given statistic. It may be used to estimate bias and variance of a statistic, to construct confidence sets and to estimate the critical values of a test statistic.

The bootstrap is attractive, because:

1. It may consistently estimate the distribution of a statistic under weaker conditions than the traditional approach (derivation of the asymptotic distribution followed by the estimation of its parameters) does.
2. It makes the derivation of the asymptotic distribution superfluous.
3. It is applicable in cases where the asymptotic distribution is untractable.
4. It has some optimality properties.

The present paper investigates the consistency properties of the bootstrap in the context of categorical data. We use an asymptotic framework, in which the model and the structure of the observed contingency tables may vary. Such "model asymptotics" or "dimension asymptotics" (since the dimension of the model is allowed to increase) have been considered before in our context by Haberman (1977a, b) and Morris (1975) and recently by Ehm (1986) and Koehler (1986). Such model asymptotics allow the examination of large, sparse contingency tables for which the requirement of standard asymptotic theory, that the minimal cell expectation be large, obviously is not fulfilled. The approach is not new; it has been used by Bickel and Freedman (1983), Huber (1973), Portnoy (1984, 1985) and Shorack (1982) for other models.

Consistency results for the bootstrap estimator of the distribution of the maximum likelihood estimator (mle) in log-linear models will be proved. In particular the phenomenon mentioned in point (1) can be observed: The bootstrap consistently estimates the distribution of linear contrasts of the mle under the weak condition

$$\kappa \cdot p^{2/3}(\log p)^{1/3} \to 0,$$

Received January 1988; revised August 1988.
[1]Work supported by the Deutsche Forschungsgemeinschaft.
AMS 1980 *subject classifications*. Primary 62G05; secondary 62H17.
Key words and phrases. Bootstrap, decomposable log-linear models, sampling models, model asymptotics, sparse contingency tables.

where p is the dimension of the log-linear model and κ is the maximal asymptotic variance of the mle's of the log expectations of the cells and κ is equal to the quantity d_t^2 of Haberman (1977b).

Under this condition the normal approximation estimator is no longer consistent since the mle can have a bias. This bias is automatically estimated by the bootstrap. The explanation for this phenomenon is that the bootstrap mimics higher order components of the distribution of the mle. A similar result has been found by Singh (1981). It is, however, not apparent how his results relate to ours.

This paper is organized in the following way: In Section 2 some preliminaries about log-linear models and sampling models for contingency tables are presented. Also the bootstrap procedure is defined in detail. The main results are presented in Section 3.

In Section 4 estimations of the Mallows distance between the true and the bootstrap distribution of the mle are derived which are conditional, given the observed contingency tables. Section 5 contains the proofs for the unconditional case. An example for the superiority of the bootstrap over the normal approximation approach and a discussion of the results are given in Section 6.

2. Preliminaries. Contingency tables will be denoted by $\mathbf{n} = (n_i\colon i \in I)$ and will be regarded as elements of $\mathbf{R}^I$. Here I is the set of cells and n_i the number of observations in cell $i \in I$. Under the Poisson sampling model the n_i are independent for $i \in I$ and n_i has a Poisson distribution with mean m_i. We shall write $\mathbf{n} \sim \mathscr{P}(\mathbf{m})$ where $\mathbf{m} = (m_i\colon i \in I)$ is the expectation vector of $\mathbf{n}$. Let $u_i = \log m_i$ and $\mathbf{u} = \log \mathbf{m} = (u_i\colon i \in I)$. Here we adapt the convention used throughout the paper that for $\mathbf{x}, \mathbf{y} \in \mathbf{R}^I$ expressions like $\mathbf{x}^2$, $\mathbf{x} \cdot \mathbf{y}$, $\log \mathbf{x}$ and $\exp \mathbf{x}$ are to be interpreted componentwise, e.g., $\mathbf{x}^2 = (x_i^2\colon i \in I)$.

For $\mathbf{u} \in \mathbf{R}^I$ let $D(\mathbf{u})$ denote the linear operator on $\mathbf{R}^I$ defined by

$$D(\mathbf{u})\mathbf{x} = (\exp(u_i)x_i\colon i \in I) = \exp(\mathbf{u}) \cdot \mathbf{x}.$$

If $\mathbf{n} \sim \mathscr{P}(\mathbf{m})$ and $\mathbf{u} = \log \mathbf{m}$, then $D(\mathbf{u})$ is the covariance operator of $\mathbf{n}$, that is, $\operatorname{cov}((\mathbf{x}, \mathbf{n}), (\mathbf{y}, \mathbf{n})) = (\mathbf{x}, D(\mathbf{u})\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbf{R}^I$ [cf. Haberman (1974) for the coordinate-free approach to contingency tables]. In this formula $(\cdot,\cdot)$ denotes the inner product

$$(\mathbf{x}, \mathbf{y}) = \sum_{i \in I} x_i y_i$$

in $\mathbf{R}^I$.

The orthogonal projection w.r.t. $(\cdot,\cdot)$ of $\mathbf{R}^I$ onto a linear subspace M of $\mathbf{R}^I$ will be denoted by π_M. We shall also make use of the weighted inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{u}} = \sum_{i \in I} x_i y_i e^{u_i} = (\mathbf{x}, D(\mathbf{u})\mathbf{y})$$

and orthogonal projections w.r.t. $\langle \cdot,\cdot \rangle_{\mathbf{u}}$ will be denoted by $\pi_M^*(\mathbf{u})$.

Under the product multinomial sampling model it is assumed that there exists a partition $\Omega = (I_1, \dots, I_K)$ of I, such that the subtables $\mathbf{n}^{(k)} = (n_i\colon i \in I_k)$ are independent for $k = 1, \dots, K$ and $\mathbf{n}^{(k)}$ has a multinomial distribution with parameters N_k, $\mathbf{p}^{(k)} = (p_i^{(k)}\colon i \in I_k)$, $p_i^{(k)} \in [0,1]$ and $\sum_{i \in I_k} p_i^{(k)} = 1$. Let $\mathbf{N} = (N_1, \dots, N_K)$ and $\mathbf{p} = (\sum_{k=1}^K 1(i \in I_k) p_i^{(k)}\colon i \in I)$. We shall use the notation

$\mathbf{n} \sim \mathscr{M}\Omega(\mathbf{N}, \mathbf{p})$, if a product multinomial model is assumed for $\mathbf{n}$.

To each sampling model there corresponds a linear subspace Z of $\mathbf{R}^I$ [cf. Haberman (1974)]. In the case of Poisson sampling $Z = \{0\}$, whereas in the case of the product multinomial sampling Z is the linear span of the K vectors $\mathbf{v}^{(k)} = (1(i \in I_k): i \in I)$. If $\mathbf{m} = E\mathbf{n}$, under each sampling model $\mathbf{n}$ satisfies the constraints $(\mathbf{n}, \mathbf{z}) = (\mathbf{m}, \mathbf{z})$ for all $\mathbf{z} \in Z$. Z is therefore called design manifold.

Under the product multinomial sampling model the covariance operator of $\mathbf{n}$ is $D(\mathbf{u})(\mathrm{Id} - \pi_Z^*(\mathbf{u}))$, if $\mathbf{u} = \log \mathbf{m}$. Since $\pi_{\{0\}}^*(\mathbf{u}) = 0$, this is also the covariance operator under the Poisson sampling model.

A log-linear model is specified by a linear subspace H of $\mathbf{R}^I$ and the requirement $\mathbf{u} \in H$. Let $\hat{\mathbf{u}}$ be the mle of $\mathbf{u}$ under the model H. It may be obtained (if it exists) by solving the equations (for $\hat{\mathbf{u}} \in H$) $\pi_H \mathbf{n} = \pi_H \hat{\mathbf{m}}$ and $\hat{\mathbf{m}} = \exp(\hat{\mathbf{u}})$. We shall make the assumption $Z \subset H$, which is discussed in Haberman (1974).

If (E, ρ) is a metric space, the Prohorov distance between two probability measures P and Q on E is the infimum of all $\varepsilon > 0$, such that for all measurable A, $P(A) \leq Q(A^\varepsilon) + \varepsilon$ and $Q(A) \leq P(A^\varepsilon) + \varepsilon$, where $A^\varepsilon = \{x \in E | \rho(x, A) \leq \varepsilon\}$. If $E = \mathbf{R}^1$ we denote the Prohorov distance by δ_1. If $E = \mathbf{R}^I$ and $\rho(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}|_1 = \sum_{i \in I} |x_i - y_i|$ is the L^1-distance, we shall write δ_I for the Prohorov distance. If X and Y have distributions P and Q we use the shorthand notation $\delta_1(X, Y)$ [or $\delta_I(X, Y)$] for $\delta_1(P, Q)$ [or $\delta_I(P, Q)$].

To estimate the distribution of $\hat{\mathbf{u}}$ the bootstrap proceeds in the following way: Given the observed table $\mathbf{n}$, let $\tilde{\mathbf{u}}$ be the value of the mle $\hat{\mathbf{u}}$. We make the distinction between $\hat{\mathbf{u}}$ (a random variable) and $\tilde{\mathbf{u}}$ (a fixed value) in order to clarify the different roles of $\hat{\mathbf{u}}$ as an estimator and its realization. Let $\tilde{\mathbf{m}} = \exp(\tilde{\mathbf{u}})$. The bootstrap estimate of the distribution of $\hat{\mathbf{u}}$ under one of the above sampling models is defined as the distribution of $\hat{\mathbf{u}}$ under the same sampling model with log-expectation $\tilde{\mathbf{u}}$. More precisely, if $\mathbf{n} \sim \mathscr{P}(\mathbf{m})$, let $\mathbf{n}^* \sim \mathscr{P}(\tilde{\mathbf{m}})$. If $\mathbf{n} \sim \mathscr{M}_\Omega(\mathbf{N}, \mathbf{p})$, let

$$\mathbf{n}^* \sim \mathscr{M}_\Omega(\mathbf{N}, \tilde{\mathbf{p}}), \quad \text{where} \quad \tilde{\mathbf{p}} = \left(\sum_{k=1}^{K} 1(i \in I_k) \frac{\tilde{m}_i}{N_k} : i \in I \right).$$

Let $\hat{\mathbf{u}}^*$ be the mle (under the log-linear model H), derived from $\mathbf{n}^*$. The distribution of $\hat{\mathbf{u}}^*$ is the bootstrap estimator of the distribution of $\hat{\mathbf{u}}$. Practically it has to be computed by a Monte Carlo simulation. The error due to this simulation will be neglected here.

It should be noted that it is necessary to use a parametric bootstrap procedure. If one imagines the table $\mathbf{n}$ as being generated by classifying observations $X_1, \ldots, X_N$, one could try to simulate the tables $\mathbf{n}^*$ by sampling with replacement from the (hypothetical) observations X_i and then classify again. However the resulting tables $\mathbf{n}^*$ have a simple $(K = 1)$ multinomial distribution with probabilities $\tilde{p}_i = n_i / N$, which is not appropriate if $\mathbf{n}$ is Poisson distributed (for example).

3. Consistency of the bootstrap estimator. We shall consider sequences $\mathbf{n}^{(t)} = (n_i^{(t)}: i \in I^{(t)})$, $t = 1, 2, \ldots,$ of contingency tables whose index set $I^{(t)}$ is allowed to depend on t. Also the assumed log-linear models $H^{(t)}$ for the log-expectation $\mathbf{u}^{(t)} = \log E\mathbf{n}^{(t)}$ and the sampling model, described by the linear manifold $Z^{(t)}$, may depend on t. We drop the index t throughout to ease notation. Convergence statements will be understood as $t \to \infty$ unless otherwise stated. Conditions for consistency involve the quantities

$$p = \dim H$$

and

$$\kappa = \max_{i \in I} \left\| \pi_H^*(\mathbf{u}) D(\mathbf{u})^{-1} \mathbf{e}^i \right\|_{\mathbf{u}}^2,$$

where $\|\mathbf{x}\|_{\mathbf{v}}^2 = \langle \mathbf{x}, \mathbf{x} \rangle_{\mathbf{v}}$ and $\mathbf{e}^i = (1(j = i)\colon\ j \in I)$. κ is the maximal $(i \in I)$ asymptotic variance of $\hat{u}_i$ under the Poisson sampling model.

The proofs are based on the examination of a second order stochastic approximation of $\hat{\mathbf{u}} - \mathbf{u}$ which is of the form

$$\hat{\mathbf{u}} - \mathbf{u} \approx \mathbf{L} - \tfrac{1}{2} \pi_H^* \mathbf{L}^2,$$

with

$$\mathbf{L} = \pi_H^*(\mathbf{u}) D(\mathbf{u})^{-1} (\mathbf{n} - \mathbf{m}),$$

$$\mathbf{L}^2 = \left(L_i^2 \colon i \in I \right)$$

[cf. Haberman (1974), equation 4.104].

Under the general asymptotic setup used here the validity of this formal approximation is difficult to prove. We therefore restrict consideration to the class of decomposable models. It seems however possible to use the methods of Ehm (1986) to extend the results to general log-linear models.

Decomposable log-linear models [see Darroch, Lauritzen and Speed (1980), Goodman (1970, 1971) and Haberman (1970, 1974)] are log-linear models for factorial tables. To define them suppose that there are factors $1, 2, \ldots, d$ and that $\bar{s}_j = \{1, \ldots, s_j\}$ is the set of levels of factor j. Then the index set I of the contingency table which summarizes the counts of the corresponding cross-classification is $I = \prod_{j=1}^d \bar{s}_j$. For a subset $E \subset \{1, \ldots, d\}$ of factors let $I(E) = \prod_{j \in E} I_j$. Furthermore let $i_E \in I(E)$ denote the index projection of $i \in I$ on $I(E)$ [i.e., $i_E = (i_j)_{j \in E}$ if $i = (i_j)_{j \in \{1, \ldots, d\}}$] and let $H(E)$ be the linear subspace of all $\mathbf{z} \in \mathbf{R}^I$ such that z_i depends only on i_E. A set $\mathfrak{A} = \{E_1, \ldots, E_T\}$ of subsets is called a generating class, if it is maximal w.r.t. inclusion. Each generating class $\mathfrak{A} = \{E_1, \ldots, E_T\}$ defines a log-linear model, namely $H(\mathfrak{A}) = H(E_1) + \cdots + H(E_T)$. $H(\mathfrak{A})$ (or $\mathfrak{A}$) is called decomposable if $T = 1$ or if there exists an ordering $\mathfrak{A} = (E_1, \ldots, E_T)$ such that for $t = 2, \ldots, T$

$$E_t \cap (E_1 \cup \cdots \cup E_{t-1}) = E_t \cap E_{r_t} =: F_t \quad \text{with } r_t < t.$$

Consistency of the bootstrap will be proven for the distribution of linear contrasts $(\mathbf{c}, \hat{\mathbf{u}})$ of the mle and for the multivariate distribution of $\hat{\mathbf{u}}$. Both are standardized in such a way that the distributions asymptotically do not degenerate. For linear contrasts we therefore assume that $\|\pi_H^*(\mathbf{u}) D(\mathbf{u})^{-1} \mathbf{c}\|_{\mathbf{u}}^2 = 1$. For contrasts orthogonal to Z this means [cf. Haberman (1977b)] that the asymptotic variance of $(\mathbf{c}, \hat{\mathbf{u}})$ is 1. Similarly (in the multivariate case) we consider $D(\mathbf{u})^{1/2}(\hat{\mathbf{u}} - \mathbf{u})$.

The following theorem contains the main result. Observe that $\mathbf{c}$ as well as the number T of sets in the generators of the log-linear models depends on t. It is assumed that either a Poisson or a product multinomial model holds.

THEOREM. *Assume that the log-linear models are decomposable with bounded* T.

(a) *If* $\|\pi_H^*(\mathbf{u})D(\mathbf{u})^{-1}\mathbf{c}\|_{\mathbf{u}}^2 = 1$ *and*

$$\text{(A)} \qquad \kappa p^{2/3}(\log p)^{1/3} \to 0$$

holds, the bootstrap consistently estimates the distribution of $(\mathbf{c}, \hat{\mathbf{u}})$, *i.e.*,

$$\delta_1((\mathbf{c}, \hat{\mathbf{u}} - \mathbf{u}), (\mathbf{c}, \hat{\mathbf{u}}^* - \tilde{\mathbf{u}})) \to 0$$

in probability. Consistency is uniform for normed contrasts $\mathbf{c}$.

(b) *If*

$$\text{(B)} \qquad |I|^2 \kappa p^2 \log p \to 0$$

holds, the bootstrap consistently estimates the distribution of $\hat{\mathbf{u}}$ *in the sense that*

$$\delta_I\left(D(\mathbf{u})^{1/2}(\hat{\mathbf{u}} - \mathbf{u}), D(\mathbf{u})^{1/2}(\hat{\mathbf{u}}^* - \tilde{\mathbf{u}})\right) \to 0$$

in probability.

We shall see later that $\kappa \geq p/N$, where N is the (expected) total number of observations. Therefore a necessary condition for (A) to hold is

$$\frac{p^{5/3}}{N}(\log p)^{1/3} \to 0.$$

We shall first examine the situation conditionally given the table $\mathbf{n}$. Then $\tilde{\mathbf{u}}$ is fixed. More precisely we assume that either (C.P) (Poisson sampling model) or (C.M) (multinomial sampling model) holds:

$$\text{(C.M)} \qquad \begin{aligned} \mathbf{n} &\sim \mathscr{M}_\Omega(\mathbf{N}, \mathbf{p}), \\ \mathbf{n}^* &\sim \mathscr{M}_\Omega(\mathbf{N}, \tilde{\mathbf{p}}). \end{aligned}$$

In this case let $\mathbf{m} = (\sum_{k=1}^K 1(i \in I_k) p_i N_k \colon i \in I)$ and $\tilde{\mathbf{m}} = (\sum_{k=1}^K 1(i \in I_k) \tilde{p}_i N_k \colon i \in I)$ be the expectations of $\mathbf{n}$ and $\mathbf{n}^*$.

$$\text{(C.P)} \qquad \begin{aligned} \mathbf{n} &\sim \mathscr{P}(\mathbf{m}), \\ \mathbf{n}^* &\sim \mathscr{P}(\tilde{\mathbf{m}}). \end{aligned}$$

Let

$$\begin{aligned} \mathbf{L} &= \pi_H^*(\mathbf{u})D(\mathbf{u})^{-1}(\mathbf{n} - \mathbf{m}), \\ \mathbf{L}^* &= \pi_H^*(\tilde{\mathbf{u}})D(\tilde{\mathbf{u}})^{-1}(\mathbf{n}^* - \tilde{\mathbf{m}}), \\ \mathbf{Q} &= \mathbf{L} - \tfrac{1}{2}\pi_H^*(\mathbf{u})\mathbf{L}^2, \\ \mathbf{Q}^* &= \mathbf{L}^* - \tfrac{1}{2}\pi_H^*(\tilde{\mathbf{u}})\mathbf{L}^{*2}, \end{aligned}$$

where $\mathbf{u} = \log \mathbf{m}$ and $\tilde{\mathbf{u}} = \log \tilde{\mathbf{m}}$.

Furthermore let

$$d_p(u, v)^p = \inf\{E|X - Y|^p \colon X \sim u, Y \sim v\}$$

be the Mallows distance [Mallows (1972)] between the probability measures u and v [see Bickel and Freedman (1981)]. Again we use the abbreviation $d_p(X, Y)$ for $d_p(u, v)$ if X and Y have distributions u and v. The L^1-norm $|\cdot|_1$ will be assumed for probability measures on $\mathbf{R}^I$.

The following proposition relates the Mallows distance between $(\mathbf{c}, \mathbf{Q})$ and $(\mathbf{c}, \mathbf{Q}^*)$ (and likewise for the multivariate distributions) to the nearness of $\mathbf{u}$ and $\tilde{\mathbf{u}}$. Technically we assume that $m_i/\tilde{m}_i$ is close to 1 uniformly in $i \in I$. Since $m_i/\tilde{m}_i - 1 = \exp(u_i - \tilde{u}_i) - 1$ this means that u_i and $\tilde{u}_i$ are close, uniformly in $i \in I$. Later on when we consider the unconditional situation the quantity ε appearing in the proposition will stand for the random variable

$$\|\mathbf{u} - \tilde{\mathbf{u}}\|_\infty = \max_{i \in I} |u_i - \tilde{u}_i|.$$

Estimating the Mallows distance is inspired by the work of Bickel and Freedman (1981).

PROPOSITION 1. *Let* $\varepsilon_0 > 0$ *and assume that* (C.M) *or* (C.P) *holds. If for some* $0 < \varepsilon \le \varepsilon_0$,

$$(*) \qquad \frac{m_i}{\tilde{m}_i} \vee \frac{\tilde{m}_i}{m_i} \le 1 + \varepsilon, \qquad i \in I,$$

then

$$d_1((\mathbf{c}, \mathbf{Q}), (\mathbf{c}, \mathbf{Q}^*)) \le \sqrt{\varepsilon}\left(C_1 + C_2\sqrt{\kappa p}\right) \cdot \left\|\pi_H^*(\mathbf{u}) D(\mathbf{u})^{-1}\mathbf{c}\right\|_{\mathbf{u}}$$

for all $\mathbf{c} \in \mathbf{R}^I$ *and also*

$$d_I\left(D(\mathbf{u})^{1/2}\mathbf{Q}, D(\mathbf{u})^{1/2}\mathbf{Q}^*\right) \le \sqrt{|I|}\sqrt{\varepsilon}\left(C_1\sqrt{p} + C_2\sqrt{\kappa p^2}\right),$$

where the constants C_1 *and* C_2 *only depend on* ε_0.

This proposition is true for arbitrary log-linear models H.

The key step in the proof of Proposition 1 is Proposition 2 which permits an estimation of the Mallows distance between the distributions of linear combinations of contingency table counts.

PROPOSITION 2. *Assume* (C.M) *or* (C.P). *Then there exists a joint distribution of* $\mathbf{n}$ *and* $\mathbf{n}^*$ *such that*

$$E\left(\sum_{i \in I} c_i((n_i - m_i) - (n_i^* - \tilde{m}_i))\right)^2 \le 2\sum_{i \in I} c_i^2 |m_i - \tilde{m}_i|.$$

Incidentally we remark that this proposition can be used to prove consistency of the bootstrap under the conventional asymptotics (which fixes the model) for a more general class of models than the log-linear model constitute. Namely, it follows directly from Proposition 2 and the definition of the Mallows distance that for a linear operator $A\colon \mathbf{R}^I \to \mathbf{R}^q$ it holds that

$$d_2^2(A(\mathbf{n}-\mathbf{m}), A(\mathbf{n}^*-\tilde{\mathbf{m}})) \le 2\sum_{l\in I}\sum_{j=1}^{q}(A^T f^j, e^l)^2 |m_l - \tilde{m}_l|.$$

Here A^T is the transpose of A and $f^j = (1(l=j)\colon l = 1,\cdots,q)$ and $e^l = (1(i=l)\colon i\in I)$ are unit vectors in $\mathbf{R}^q$ and $\mathbf{R}^I$.

Assume that there is a parametric model for the cell expectations $m_i = m_i(\theta)$, $\theta\in\Theta$. Frequently [cf. Bishop, Fienberg, and Holland (1975), Theorem 14.8-3] an estimator $\hat{\theta}$ of θ has a stochastic expansion

$$\sqrt{N}(\hat{\theta}-\theta) = A\frac{1}{\sqrt{N}}(\mathbf{n}-\mathbf{m}(\theta)) + o_p(1),$$

which is valid locally uniformly. Then

$$d_2^2\left(A\frac{1}{\sqrt{N}}(\mathbf{n}-\mathbf{m}(\theta)), A\frac{1}{\sqrt{N}}(\mathbf{n}^*-m(\hat{\theta}))\right) \le 2\max_{i\in I}\left|\frac{m_i(\theta)-m_i(\hat{\theta})}{N}\right|\sum_{l\in I}\sum_{j=1}^{q}(A^T f^j, e^l)^2.$$

So the bootstrap is consistent if the cell frequencies $p_i(\theta) = m_i(\theta)/N$ are estimated consistently by $p_i(\hat{\theta})$.

To apply Proposition 1 it is necessary to determine the rate of convergence of $\|\hat{\mathbf{u}}-\mathbf{u}\|_\infty$ to zero. We shall prove the following.

PROPOSITION 3. *Assume that the log-linear models are decomposable with bounded T. Then if* $\kappa\log p\to 0$,

$$\|\hat{\mathbf{u}}-\mathbf{u}\|_\infty = O_p\left(\sqrt{\kappa\log p}\right).$$

The following proposition shows that it is enough (under the conditions of Theorem 1) to consider the approximation of $\hat{\mathbf{u}}-\mathbf{u}$ by $\mathbf{Q}$. The distribution of $\mathbf{n}$ will be denoted by $P_{\mathbf{u}}$ and the conditional distribution of $\mathbf{n}^*$ (given $\mathbf{n}$) by $P^*_{\hat{\mathbf{u}}}$. Unconditionally $P^*_{\hat{\mathbf{u}}}$ is a random distribution.

PROPOSITION 4. *Assume that the log-linear models are decomposable with bounded T.*

(a) *If* $\|\pi_H^*(\mathbf{u})D(\mathbf{u})^{-1}\mathbf{c}\|_{\mathbf{u}} = 1$ *and* (A) *holds, then for all* $\varepsilon>0$,

$$P_{\mathbf{u}}\left(|(\mathbf{c},\hat{\mathbf{u}}-\mathbf{u}-\mathbf{Q})|>\varepsilon\right)\to 0,$$

$$P^*_{\hat{\mathbf{u}}}\left(|(\mathbf{c},\hat{\mathbf{u}}^*-\hat{\mathbf{u}}-\mathbf{Q}^*)|>\varepsilon\right)\to 0 \quad \text{in } P_{\mathbf{u}}\text{-probability}.$$

Convergences are uniform for normed contrasts $\mathbf{c}$.

(b) *If* (B) *holds, then for all* $\varepsilon>0$,

$$P_{\mathbf{u}}\left(\left|D(\mathbf{u})^{1/2}(\hat{\mathbf{u}}-\mathbf{u}-\mathbf{Q})\right|_1>\varepsilon\right)\to 0,$$

$$P^*_{\hat{\mathbf{u}}}\left(\left|D(\mathbf{u})^{1/2}(\hat{\mathbf{u}}^*-\hat{\mathbf{u}}-\mathbf{Q}^*)\right|_1>\varepsilon\right)\to 0 \quad \text{in } \mathbf{P}_{\mathbf{u}}\text{-probability}.$$

4. Proofs—conditional distributions. In this section we prove Propositions 1 and 2. These give estimations for fixed $\mathbf{u}$ and $\tilde{\mathbf{u}}$ and may be thought of as describing the situation for a given (observed) contingency table $\mathbf{n}$.

PROOF OF PROPOSITION 2. Assume that (C.P) holds for $\mathbf{n} = (n_i: i \in I)$ and $\mathbf{n}^* = (n_i^*: i \in I)$. Since the n_i are independent and the same is true for the n_i^*, we can construct a joint distribution of $\mathbf{n}$ and $\mathbf{n}^*$ by specifying a joint distribution of n_i and n_i^* for each $i \in I$ and letting the pairs (n_i, n_i^*) be independent. Then we have

$$E\Big(\sum_{i\in I} c_i((n_i - m_i) - (n_i^* - \tilde{m}_i))\Big)^2 = \sum_{i\in I} c_i^2 E((n_i - m_i) - (n_i^* - \tilde{m}_i))^2.$$

Therefore it is sufficient to prove that for two univariate random variables $n \sim \mathscr{P}(m)$ and $n^* \sim \mathscr{P}(\tilde{m})$ there exist a joint distribution with

$$E((n - m) - (n^* - \tilde{m}))^2 \le 2|m - \tilde{m}|.$$

W.l.o.g. assume that $m < \tilde{m}$. For some $N \in \{1, 2, \dots\}$ which will be chosen later let $X \sim \mathscr{P}(p)$ and $X^* \sim \mathscr{P}(\tilde{p})$ with $p = m/N$ and $\tilde{p} = \tilde{m}/N$.

A joint distribution of X and X^* can (and shall) be constructed such that $P(X = j, X^* = j) = P(X = j) \wedge P(X^* = j)$ for all $j = 0, 1, 2, \dots$ (in other words, the event $\{X = X^*\}$ has maximal probability). Since $P(X = i)/P(X^* = i) = \gamma^i e^{\delta/N}$ with $\gamma = m/\tilde{m} < 1$ and $\delta = \tilde{m} - m > 0$, by choosing N large enough one can achieve

$$P(X = i, X^* = i) = P(X = i) \wedge P(X^* = i) = P(X = i), \qquad i > 0. \tag{4.1}$$

Now let $(X^{(l)}, X^{*(l)})$ $(l = 1, \dots, N)$ be independent copies of (X, X^*) and let $(n, n^*) = \sum_{l=1}^N (X^{(l)}, X^{*(l)})$. Then $n \sim \mathscr{P}(m)$ and $n^* \sim \mathscr{P}(\tilde{m})$. We can estimate

$$\begin{aligned}
&E((n - m) - (n^* - \tilde{m}))^2 \\
&\quad = NE((X - p) - (X^* - \tilde{p}))^2 \le NE(X - X^*)^2 \\
&\quad = N\sum_{i\ne j} (i - j)^2 P(X = i, X^* = j) \\
&\quad \le 2N\Big\{\sum_{i\ne j} i^2 P(X = i, X^* = j) + \sum_{i\ne j} j^2 P(X = i, X^* = j)\Big\} \\
&\quad = 2N\Big\{\sum_{i=0}^{\infty} i^2 P(X = i, X^* \ne i) + \sum_{i=0}^{\infty} i^2 P(X^* = i, X \ne i)\Big\} \\
&\quad = 2N\Big\{\sum_{i=0}^{\infty} i^2 [P(X = i) - P(X = i, X^* = i) + P(X^* = i) \\
&\qquad\qquad - P(X = i, X^* = i)]\Big\} \\
&\quad = 2N\Big\{\sum_{i=1}^{\infty} i^2 (P(X^* = i) - P(X = i))\Big\}
\end{aligned}$$

by (4.1). The last expression is equal to

$$2N\{E(X^*)^2 - EX^2\} = 2|\tilde{m} - m|(1 + \tilde{p} + p).$$

Now let $N \to \infty$. The resulting sequence of joint distributions of n and n^* is tight. Proposition 2 is proved therefore for the Poisson sampling model.

Now assume that (C.M) holds. The case $K > 1$ can easily be reduced to the case $K = 1$ by independence as above. If $K = 1$, $\mathbf{n} \sim \mathscr{M}(N, \mathbf{p})$ and $\mathbf{n}^* \sim \mathscr{M}(N, \tilde{\mathbf{p}})$, write $\mathbf{n} = \sum_{l=1}^N \mathbf{X}^{(l)}$, $\mathbf{n}^* = \sum_{l=1}^N \mathbf{X}^{*(l)}$ where $\mathbf{X}^{(l)} \sim \mathscr{M}(1, \mathbf{p})$ and $\mathbf{X}^{*(l)} \sim \mathscr{M}(1, \tilde{\mathbf{p}})$, with independent $\mathbf{X}^{(l)}$ and independent $\mathbf{X}^{*(l)}$. A joint distribution of $\mathbf{X}^{(l)}$ and $\mathbf{X}^{*(l)}$ can then be constructed which satisfies

$$P\left(X_i^{(l)} = 1,\ X_i^{*(l)} = 1\right) = p_i \wedge \tilde{p}_i, \qquad i \in I.$$

Similar estimates as above show that Proposition 2 is also true under the multinomial sampling model. □

The proof of Proposition 1 will be prepared by a series of lemmas. We still consider $\mathbf{m}$ and $\tilde{\mathbf{m}}$ as fixed. $D(\mathbf{u})$ and $D(\tilde{\mathbf{u}})$ will be abbreviated by D and $\tilde{D}$. Also we write $\pi_M^* = \pi_M^*(\mathbf{u})$ and $\tilde{\pi}_M^* = \pi_M^*(\tilde{\mathbf{u}})$. As in Section 2, $\mathbf{L} = \pi_H^* D^{-1}(\mathbf{n} - \mathbf{m})$, $\mathbf{L}^* = \tilde{\pi}_H^* \tilde{D}^{-1}(\mathbf{n}^* - \tilde{\mathbf{m}})$, $\mathbf{Q} = \mathbf{L} - \frac{1}{2}\pi_H^* \mathbf{L}^2$ and $\mathbf{Q}^* = \mathbf{L}^* - \frac{1}{2}\tilde{\pi}_H^* \mathbf{L}^{*2}$ are the first and second order approximations to the centered mle.

LEMMA 1. (i) *For* $\mathbf{v} \in \mathbf{R}^I$ *we have*

$$\pi_M^*(\mathbf{v}) D(\mathbf{v})^{-1} = \pi_M^*(\mathbf{v}) D(\mathbf{v})^{-1} \pi_M.$$

(ii) *For* $\mathbf{v} \in \mathbf{R}^I$ *the linear operators* $\pi_M^*(\mathbf{v})D(\mathbf{v})^{-1}$ *and* $\pi_M D(\mathbf{v})$ *are mutually inverse on* M.

PROOF. Part (i) holds since $(\mathbf{X}, \mathbf{m}) = 0$ for all $\mathbf{m} \in M$ is equivalent to $\langle D(\mathbf{v})^{-1}\mathbf{X}, \mathbf{m}\rangle_{\mathbf{v}} = 0$ for all $\mathbf{m} \in M$. Part (ii) follows since (for $\mathbf{m} \in M$) $\pi_M^*(\mathbf{v})D(\mathbf{v})^{-1}\pi_M D(\mathbf{v})\mathbf{m} = \pi_M^*(\mathbf{v})D(\mathbf{v})^{-1}D(\mathbf{v})\mathbf{m} = \mathbf{m}$. □

LEMMA 2. *The operator* $\pi_M^*(\mathbf{v})D(\mathbf{v})^{-1}$ *on* $\mathbf{R}^I$ *is symmetric w.r.t. the inner product* $(\cdot, \cdot)$.

PROOF. This follows immediately from the fact that $\pi_M^*(\mathbf{v})$ is symmetric w.r.t. $\langle \cdot, \cdot \rangle_{\mathbf{v}}$, since

$$\left(\pi_M^*(\mathbf{v})D(\mathbf{v})^{-1}\mathbf{x}, \mathbf{y}\right) = \left\langle \pi_M^*(\mathbf{v})D(\mathbf{v})^{-1}\mathbf{x}, D(\mathbf{v})^{-1}\mathbf{y}\right\rangle_{\mathbf{v}}. \qquad □$$

Before stating the next lemma we introduce the operator norm $\|A\|_{\mathbf{v}} = \sup_{\|\mathbf{x}\|_{\mathbf{v}} \le 1}\|A\mathbf{x}\|_{\mathbf{v}}$. Since $\pi_M^*(\mathbf{v})$ is an orthogonal projection w.r.t. $\langle \cdot, \cdot \rangle_{\mathbf{v}}$, $\|\pi_M^*(\mathbf{v})\|_{\mathbf{v}} \le 1$. The next lemma estimates $\|\pi_M^*(\tilde{\mathbf{u}})\|_{\mathbf{u}}$ for $\tilde{\mathbf{u}}$ close to $\mathbf{u}$.

LEMMA 3. *Assume that for some* $\varepsilon \ge 0$,

$$\frac{m_i}{\tilde{m}_i} \vee \frac{\tilde{m}_i}{m_i} \le 1 + \varepsilon, \qquad i \in I, \tag{4.2}$$

holds. Then we have

$$\|\pi_M^*(\tilde{\mathbf{u}})\|_{\mathbf{u}} \le 1 + \varepsilon.$$

PROOF. It follows from (4.2) that $\|\mathbf{x}\|_{\mathbf{u}} \le \sqrt{1+\varepsilon}\,\|\mathbf{x}\|_{\tilde{\mathbf{u}}}$ and vice versa. Therefore

$$\|\pi_H^*(\tilde{\mathbf{u}})\|_{\mathbf{u}} \le \sup_{\|\mathbf{x}\|_{\mathbf{u}} \le 1} \sqrt{1+\varepsilon}\,\|\pi_H^*(\tilde{\mathbf{u}})\|_{\tilde{\mathbf{u}}}\|\mathbf{x}\|_{\tilde{\mathbf{u}}} \le 1 + \varepsilon. \qquad \square$$

LEMMA 4. *Let* $\mathbf{c} = (c_i\colon i \in I)$ *be in* $\mathbf{R}^I$ *and assume that* (4.2) *holds. Then*

$$\left\|\left(\pi_H^* D^{-1} - \tilde{\pi}_H^* \tilde{D}^{-1}\right)\mathbf{c}\right\|_{\mathbf{u}} \le \varepsilon(1+\varepsilon)\|\pi_H^* D^{-1}\mathbf{c}\|_{\mathbf{u}}.$$

PROOF. According to Lemma 1,

$$\tilde{\pi}_H^*\tilde{D}^{-1} = \tilde{\pi}_H^*\tilde{D}^{-1}\pi_H = \tilde{\pi}_H^*\tilde{D}^{-1}\pi_H D \pi_H^* D^{-1}\pi_H = \tilde{\pi}_H^*\tilde{D}^{-1}D\pi_H^*D^{-1}.$$

Therefore,

$$\begin{aligned}\left\|\left(\pi_H^* D^{-1} - \tilde{\pi}_H^*\tilde{D}^{-1}\right)\mathbf{c}\right\|_{\mathbf{u}} &= \left\|\tilde{\pi}_H^*(\mathrm{Id} - \tilde{D}^{-1}D)\pi_H^* D^{-1}\mathbf{c}\right\|_{\mathbf{u}}\\ &\le \|\tilde{\pi}_H^*\|_{\mathbf{u}}\left\|(\mathrm{Id} - \tilde{D}^{-1}D)\pi_H^*D^{-1}\mathbf{c}\right\|_{\mathbf{u}}.\end{aligned}$$

By Lemma 3 $\|\tilde{\pi}_H^*\|_{\mathbf{u}} \le 1 + \varepsilon$. Also for $\mathbf{z} \in \mathbf{R}^I$,

$$\left\|(\mathrm{Id} - \tilde{D}^{-1}D)\mathbf{z}\right\|_{\mathbf{u}}^2 = \sum_{i \in I} m_i(1 - m_i/\tilde{m}_i)^2 z_i^2 \le \varepsilon^2\|\mathbf{z}\|_{\mathbf{u}}^2.$$

Therefore the assertion of the lemma is true. $\square$

The next lemma contains the estimation of the Mallows distance between the linear approximations of the mle under the true and under the bootstrap distribution.

LEMMA 5. *Let* $\mathbf{c} = (c_i\colon i \in I)$ *be in* $\mathbf{R}^I$ *and assume that* (4.2) *holds. For the joint distribution of* $\mathbf{n}$ *and* $\mathbf{n}^*$ *constructed in the proof of Proposition* 2

$$E(\mathbf{c}, \mathbf{L} - \mathbf{L}^*)^2 \le \varepsilon\left(4 + 2\varepsilon(1+\varepsilon)^4\right)\|\pi_H^* D^{-1}\mathbf{c}\|_{\mathbf{u}}^2.$$

PROOF. We have

$$\begin{aligned}E(\mathbf{c}, \mathbf{L} - \mathbf{L}^*)^2 &= E\Big(\big(\mathbf{c}, \pi_H^* D^{-1}[(\mathbf{n} - \mathbf{m}) - (\mathbf{n}^* - \tilde{\mathbf{m}})]\big)\\ &\qquad + \big(\mathbf{c}, \left(\pi_H^* D^{-1} - \tilde{\pi}_H^*\tilde{D}^{-1}\right)(\mathbf{n}^* - \tilde{\mathbf{m}})\big)\Big)^2\\ &\le 2\Big\{E\big(\mathbf{c}, \pi_H^* D^{-1}[(\mathbf{n} - \mathbf{m}) - (\mathbf{n}^* - \tilde{\mathbf{m}})]\big)^2\\ &\qquad + E\big(\mathbf{c}, \left(\pi_H^* D^{-1} - \tilde{\pi}_H^*\tilde{D}^{-1}\right)(\mathbf{n}^* - \tilde{\mathbf{m}})\big)^2\Big\}\\ &= 2\Big\{E\big(\pi_H^* D^{-1}\mathbf{c}, (\mathbf{n} - \mathbf{m}) - (\mathbf{n}^* - \tilde{\mathbf{m}})\big)^2\\ &\qquad + E(\Gamma\mathbf{c}, \mathbf{n}^* - \tilde{\mathbf{m}})^2\Big\},\end{aligned}$$

where $\Gamma = \pi_H^* D^{-1} - \tilde{\pi}_H^* \tilde{D}^{-1}$. Here we applied Lemma 2 which is also valid with $\pi_H^* D^{-1}$ replaced by Γ. Let $\mathbf{c}' = \pi_H^* D^{-1}\mathbf{c}$. By Proposition 2 the first summand is less than or equal to

$$2\sum_{i\in I}(c_i')^2|m_i - \tilde{m}_i| \le 2\max_{i\in I}\left|\frac{m_i - \tilde{m}_i}{m_i}\right| \|\mathbf{c}'\|_{\mathbf{u}}^2 \le 2\varepsilon\|\pi_H^* D^{-1}\mathbf{c}\|_{\mathbf{u}}^2$$

by (4.2). The second summand is equal to

$$\begin{aligned} 2\operatorname{Var}(\Gamma\mathbf{c}, \mathbf{n}^*) &= 2\big(\Gamma\mathbf{c}, \tilde{D}(\mathrm{Id} - \tilde{\pi}_Z^*)\Gamma\mathbf{c}\big) \\ &= 2\big\langle \Gamma\mathbf{c}, D^{-1}\tilde{D}(\mathrm{Id} - \tilde{\pi}_Z^*)\Gamma\mathbf{c}\big\rangle_{\mathbf{u}} \\ &\le \|D^{-1}\tilde{D}\|_{\mathbf{u}}\|\mathrm{Id} - \tilde{\pi}_Z^*\|_{\mathbf{u}}\|\Gamma c\|_{\mathbf{u}}^2. \end{aligned}$$

It is easy to se that $\|D^{-1}\tilde{D}\|_{\mathbf{u}} \le 1 + \varepsilon$. The result follows now from Lemmas 3 and 4. □

PROOF OF PROPOSITION 1. Assume that $\mathbf{n}$ and $\mathbf{n}^*$ have the joint distribution constructed in Proposition 2. Then

$$\begin{aligned} d_1((\mathbf{c}, \mathbf{Q}), (\mathbf{c}, \mathbf{Q}^*)) &\le E\big|\big(\mathbf{c}, \mathbf{L} - \tfrac{1}{2}\pi_H^*\mathbf{L}^2\big) - \big(\mathbf{c}, \mathbf{L}^* - \tfrac{1}{2}\tilde{\pi}_H^*\mathbf{L}^{*2}\big)\big| \\ &\le \sqrt{E(\mathbf{c}, \mathbf{L} - \mathbf{L}^*)^2} + \tfrac{1}{2}E\big|\big(\mathbf{c}, \pi_H^*\mathbf{L}^2 - \tilde{\pi}_H^*\mathbf{L}^{*2}\big)\big| \\ &\le \sqrt{\varepsilon}\, C_1\|\pi_H^* D^{-1}\mathbf{c}\|_{\mathbf{u}} + \tfrac{1}{2}E\big|\big(\mathbf{c}, \pi_H^*\mathbf{L}^2 - \tilde{\pi}_H^*\mathbf{L}^{*2}\big)\big|, \end{aligned}$$

where $C_1 = \sqrt{4 + 2\varepsilon_0(1 + \varepsilon_0)^4}$ (use the Schwarz inequality and Lemma 5).

It remains to estimate $E|(\mathbf{c}, \pi_H^*\mathbf{L}^2 - \tilde{\pi}_H^*\mathbf{L}^{*2})|$, which according to Lemma 2 can be written as

$$\begin{aligned} &E\big|\big(\pi_H^* D^{-1}\mathbf{c}, D\mathbf{L}^2\big) - \big(\tilde{\pi}_H^*\tilde{D}^{-1}\mathbf{c}, \tilde{D}\mathbf{L}^{*2}\big)\big| \\ &\quad = E\big|\langle \pi_H^* D^{-1}\mathbf{c}, \mathbf{L}^2\rangle_{\mathbf{u}} - \langle \tilde{\pi}_H^*\tilde{D}^{-1}\mathbf{c}D^{-1}, \tilde{D}\mathbf{L}^{*2}\rangle_{\mathbf{u}}\big| \\ &\quad \le E|\eta_1| + E|\eta_2| + E|\eta_3|, \end{aligned}$$

where

$$\eta_1 = \langle \mathbf{c}', \mathbf{L}^2 - \mathbf{L}^{*2}\rangle_{\mathbf{u}}, \qquad \eta_2 = \langle \mathbf{c}', \mathbf{L}^{*2} - D^{-1}\tilde{D}\mathbf{L}^{*2}\rangle_{\mathbf{u}},$$
$$\eta_3 = \langle \mathbf{c}' - \mathbf{c}'', D^{-1}\tilde{D}\mathbf{L}^{*2}\rangle_{\mathbf{u}}, \qquad \mathbf{c}' = \pi_H^* D^{-1}\mathbf{c}, \mathbf{c}'' = \tilde{\pi}_H^*\tilde{D}^{-1}\mathbf{c}.$$

We first estimate the main term $E|\eta_1|$. Repeated application of the Schwarz inequality yields

$$\begin{aligned} E|\eta_1| &= E\left|\sum_{i\in I} m_i c_i'\big(L_i^2 - L_i^{*2}\big)\right| \\ &\le E\sum_{i\in I} m_i|c_i'|\,|L_i - L_i^*|\,|L_i + L_i^*| \\ &\le \sqrt{\sum_{i\in I} m_i|c_i'|^2 \sum_{i\in I} m_i E(L_i - L_i^*)^2 E(L_i + L_i^*)^2} \\ &= \|\pi_H^* D^{-1}\mathbf{c}\|_{\mathbf{u}}\sqrt{\sum_{i\in I} m_i E(L_i - L_i^*)^2 E(L_i + L_i^*)^2}. \end{aligned}$$

By Lemma 5, $E(L_i - L_i^*)^2 = E(\mathbf{e}^i, \mathbf{L} - \mathbf{L}^*)^2 \le \varepsilon C_1^2 \kappa_i$ and $\kappa_i = \|\pi_H^* D^{-1}\mathbf{e}^i\|_{\mathbf{u}}$. Furthermore $E(L_i + L_i^*)^2 \le 2(\operatorname{Var} L_i + \operatorname{Var} L_i^*)$.

Since $D(\mathrm{Id} - \pi_Z^*)$ is the covariance operator of $\mathbf{n}$,

$$\operatorname{Var} L_i = \left\|(\mathrm{Id} - \pi_Z^*)\pi_H^* D^{-1}\mathbf{e}^i\right\|_{\mathbf{u}}^2 \le \|\pi_H^* D^{-1}\mathbf{e}^i\|_{\mathbf{u}}^2 = \kappa_i.$$

Similarly,

$$\begin{aligned}\operatorname{Var} L_i^* &\le \|\tilde{\pi}_H^* \tilde{D}^{-1}\mathbf{e}^i\|_{\tilde{\mathbf{u}}}^2 \\ &\le \|D^{-1/2}\tilde{D}^{1/2}\|_{\mathbf{u}}^2\left(\|\tilde{\pi}_H^*\tilde{D}^{-1}\mathbf{e}^i - \pi_H^* D^{-1}\mathbf{e}^i\|_{\mathbf{u}}^2 + \|\pi_H^* D^{-1}\mathbf{e}^i\|_{\mathbf{u}}^2\right) \\ &\le 2(1+\varepsilon_0)\left(\varepsilon_0^2(1+\varepsilon_0)^2 + 1\right)\kappa_i\end{aligned}$$

(Lemma 4).

Combining these estimates yields

$$\begin{aligned}E|\eta_1| &\le C'_2\sqrt{\sum_{i\in I} m_i \varepsilon \kappa_i^2}\,\|\pi_H^* D^{-1}\mathbf{c}\|_{\mathbf{u}} \\ &\le C'_2\sqrt{\varepsilon\kappa \sum m_i \kappa_i}\,\|\pi_H^* D^{-1}\mathbf{c}\|_{\mathbf{u}} \quad \text{for some constant } C_2'.\end{aligned}$$

As was also shown by Haberman (1977b), Lemma 2,

$$\begin{aligned}\sum_{i\in I} m_i\kappa_i &= \sum_{i\in I} m_i\langle \pi_H^* D^{-1}\mathbf{e}^i, D^{-1}\mathbf{e}^i\rangle_{\mathbf{u}} \\ &= \sum_{i\in I}\langle \pi_H^* D^{-1}\mathbf{e}^i, \mathbf{e}^i\rangle_{\mathbf{u}} = \sum_{i\in I}\left(\mathbf{e}^i, \pi_H^*\mathbf{e}^i\right) \qquad (4.3) \\ &= \operatorname{trace}\pi_H^* = \dim H = p.\end{aligned}$$

Therefore

$$E|\eta_1| \le C_2'\sqrt{\varepsilon\kappa p}\,\|\pi_H^* D^{-1}\mathbf{c}\|_{\mathbf{u}}.$$

Similar estimates yield

$$\begin{aligned}E|\eta_2| &= E\left|\left\langle \mathbf{c}', (\mathrm{Id} - D^{-1}\tilde{D})\mathbf{L}^{*2}\right\rangle_{\mathbf{u}}\right| \\ &\le \sum_{i\in I} m_i|c_i'|\left|1 - \frac{\tilde{m}_i}{m_i}\right| EL_i^{*2} \\ &\le \sum_{i\in I} m_i|c_i'|\varepsilon C_2''\kappa_i \\ &\le \varepsilon C_2''\sqrt{\kappa p}\,\|\pi_H^* D^{-1}\mathbf{c}\|_{\mathbf{u}} \quad \text{for some constant } C_2''.\end{aligned}$$

Finally

$$\begin{aligned}E|\eta_3| &\le \sum_{i\in I} m_i|c_i' - c_i''|\left|\frac{\tilde{m}_i}{m_i}\right| EL_i^{*2} \\ &\le \|\mathbf{c}' - \mathbf{c}''\|_{\mathbf{u}}\sqrt{\sum_{i\in I} m_i\left(EL_i^{*2}\right)^2}\,(1+\varepsilon_0).\end{aligned}$$

Now use Lemma 4 and obtain

$$E|\eta_3| \le \varepsilon C_2''' \sqrt{\kappa p}\, \|\pi_H^* D^{-1}\mathbf{c}\|_{\mathbf{u}} \quad \text{for some constant } C_2'''.$$

We have therefore shown that

$$E\big|\big(\mathbf{c}, \pi_H^* \mathbf{L}^2 - \tilde{\pi}_H^* \mathbf{L}^{*2}\big)\big| \le C_2\sqrt{\varepsilon\kappa p}\, \|\pi_H^* D^{-1}\mathbf{c}\|_{\mathbf{u}} \quad \text{for some constant } C_2.$$

The first assertion of Proposition 1 is therefore proved. For the second assertion observe that

$$\begin{aligned} d_1(D^{1/2}\mathbf{Q}, D^{1/2}\mathbf{Q}^*) &\le E\big|D^{1/2}(\mathbf{Q} - \mathbf{Q}^*)\big|_1 \\ &= \sum_{i\in I} E\big|\big(\mathbf{e}^i, D^{1/2}(\mathbf{Q} - \mathbf{Q}^*)\big)\big| \\ &= \sum_{i\in I} E\big|\big(D^{1/2}\mathbf{e}^i, \mathbf{Q} - \mathbf{Q}^*\big)\big|. \end{aligned}$$

Now apply the just proved assertion with $\mathbf{c} = D^{1/2}\mathbf{e}^i$ and obtain

$$\begin{aligned} d_I(D^{1/2}\mathbf{Q}, D^{1/2}\mathbf{Q}^*) &\le \sqrt{\varepsilon}\big(C_1 + C_2\sqrt{\kappa p}\big) \sum_{i\in I} \|\pi_H^* D^{-1} D^{1/2}\mathbf{e}^i\|_{\mathbf{u}} \\ &\le \sqrt{\varepsilon}\big(C_1 + C_2\sqrt{\kappa p}\big)\sqrt{|I|}\, \sqrt{\sum_{i\in I} \|\pi_H^* D^{-1} D^{1/2}\mathbf{e}^i\|_{\mathbf{u}}^2} \end{aligned}$$

according to the Schwarz inequality. Now note that $\sum_{i\in I}\|\pi_H^* D^{-1}D^{1/2}\mathbf{e}^i\|_{\mathbf{u}}^2 = p$ by (4.3). This proves the second assertion of Proposition 1. □

5. Proofs—decomposable models. In this section $\tilde{\mathbf{u}}$ will be the random bootstrap parameter which was taken as the mle $\hat{\mathbf{u}}$. We will first obtain information on the convergence rate of $\hat{\mathbf{u}}$ and then apply Proposition 1. This will be facilitated by confining to decomposable log-linear models H. Heavy use will be made of the fact that in these models the mle has an explicit representation.

We shall state this first. Details of the following computations can be found in Andersen (1974), Darroch, Lauritzen and Speed (1980) and Haberman (1974). Assume that factors $1,\dots,d$ are given and that $H = H(\mathfrak{A})$ is decomposable with (ordered) generator $\mathfrak{A} = (E_1,\dots,E_T)$, $E_t \subset \{1,\dots,d\}$. For $\mathbf{x} \in \mathbf{R}^I$ and $E_t \subset \{1,\dots,d\}$ let $\mathbf{x}^E = (\sum_{k\in I,\, k_E = i_E} x_k\colon i \in I)$ be the E-marginal of $\mathbf{x}$ (recall that k_E and i_E are the index projections of k, $i \in I = \prod_{j=1}^d I_j$, Section 2). Clearly $\mathbf{x}^E \in H(E)$ and one can easily prove that $\pi_{H(E)}\mathbf{x} = f(E)\mathbf{x}^E$ and $1/f(E) = \prod_{j\in\{1,\dots,d\}\setminus E} s_j$. Now let $\mathbf{x} \in R^I$, $\mathbf{u} = \log \mathbf{m} \in \mathbf{R}^I$ and $\mathbf{z} = \pi_H^*(\mathbf{u})D(\mathbf{u})^{-1}\mathbf{x}$. Then by Lemma 1

$$\begin{aligned} \pi_{H(E)}\mathbf{x} &= \pi_{H(E)}D(\mathbf{u})\pi_{H(E)}^*(\mathbf{u})D(\mathbf{u})^{-1}\pi_{H(E)}\mathbf{x} \\ &= \pi_{H(E)}D(\mathbf{u})\mathbf{z}, \end{aligned}$$

and therefore $f(E)\mathbf{x}^E = f(E)(D(\mathbf{u})\mathbf{z})^E = f(E)\mathbf{z}\mathbf{m}^E$ [since $\mathbf{z} \in H(E)$].

It follows that $\pi_H^*(\mathbf{u})D(\mathbf{u})^{-1}\mathbf{x} = \mathbf{x}^E/\mathbf{m}^E$. Let $\mathbf{u} = \log \mathbf{m} \in H(\mathfrak{A})$ and $\hat{\mathbf{u}} = \log \hat{\mathbf{m}}$ be the mle of $\mathbf{u}$. It was shown by Darroch, Lauritzen and Speed (1980), pages 529–531 that the mle exists iff $n_i^{E_t} > 0$ for all $i \in I$, $t = 1, \ldots, T$, and that

$$m_i = \frac{\prod_{t=1}^T m_i^{E_t}}{\prod_{t=2}^T m_i^{F_t}}, \qquad \hat{m}_i = \frac{\prod_{t=1}^T n_i^{E_t}}{\prod_{t=2} n_i^{F_t}}, \qquad i \in I,$$

and therefore

$$\begin{aligned} \hat{u}_i - u_i &= \log\left(\frac{\hat{m}_i}{m_i}\right) \\ &= \sum_{t=1}^T \log\left(1 + \frac{n_i^{E_t} - m_i^{E_t}}{m^{E_t}}\right) - \sum_{t=2}^T \log\left(1 + \frac{n_i^{F_t} - m_i^{F_t}}{m^{F_t}}\right). \end{aligned} \tag{5.1}$$

The following lemma summarizes results hidden in the literature [Haberman (1974), Chapter 5, especially page 208]. It gives explicit formulas for the quantities **L**, **Q** and κ_i. We present it with only a brief indication of how a structural proof can be performed.

Lemma 6. *Assume that $H = H(\mathfrak{A})$ is decomposable with generator $\mathfrak{A} = (E_1, \ldots, E_T)$. Then for $\mathbf{u} = \log \mathbf{m} \in H$:*

(i) $$\pi_H^*(\mathbf{u}) D(\mathbf{u})^{-1}\mathbf{x} = \sum_{t=1}^T \frac{\mathbf{x}^{E_t}}{\mathbf{m}^{E_t}} - \sum_{t=2}^T \frac{\mathbf{x}^{F_t}}{\mathbf{m}^{F_t}}, \qquad \mathbf{x} \in \mathbf{R}^I.$$

(ii) $$\mathbf{L} = \sum_{t=1}^T \frac{\mathbf{n}^{E_t} - \mathbf{m}^{E_t}}{\mathbf{m}^{E_t}} - \sum_{t=2}^T \frac{\mathbf{n}^{F_t} - \mathbf{m}^{F_t}}{\mathbf{m}^{F_t}}.$$

(iii) $$\begin{aligned} \mathbf{Q} = &\sum_{t=1}^T \left\{ \frac{\mathbf{n}^{E_t} - \mathbf{m}^{E_t}}{\mathbf{m}^{E_t}} - \frac{1}{2}\left(\frac{\mathbf{n}^{E_t} - \mathbf{m}^{E_t}}{\mathbf{m}^{E_t}}\right)^2 \right\} \\ &- \sum_{t=2}^T \left\{ \frac{\mathbf{n}^{F_t} - \mathbf{m}^{F_t}}{\mathbf{m}^{F_t}} - \frac{1}{2}\left(\frac{\mathbf{n}^{F_t} - \mathbf{m}^{F_t}}{\mathbf{m}^{F_t}}\right)^2 \right\}. \end{aligned}$$

(iv) $$\kappa_i = \sum_{t=1}^T \frac{1}{m_i^{E_t}} - \sum_{t=2}^T \frac{1}{m_i^{F_t}}.$$

Proof. With the methods of Haberman (1974), it can be shown by induction on T that

$$H(\mathfrak{A}) = H(E_1) \oplus \bigoplus_{t=2}^T (H(E_t) \ominus H(F_t)),$$

where $M_1 \oplus M_2$ denotes the direct sum of subspaces M_i of $\mathbf{R}^I$ that are orthogonal with respect to $\langle \cdot, \cdot \rangle_{\mathbf{u}}$ and $M_1 \ominus M_2$ is the respective orthogonal complement of M_2 in M_1. Then

$$\pi_{H(\mathfrak{A})}^*(\mathbf{u}) D(\mathbf{u})^{-1}\mathbf{x} = \sum_{t=1}^T \pi_{H(E_t)}^*(\mathbf{u}) D(\mathbf{u})^{-1}\mathbf{x} - \sum_{t=2}^T \pi_{H(F_t)}^*(\mathbf{u}) D(\mathbf{u})^{-1}\mathbf{x},$$

from which (i) follows. Parts (ii) and (iv) follow from (i) by letting $\mathbf{x} = \mathbf{n} - \mathbf{m}$ and $\mathbf{x} = \mathbf{e}^i$. Again by induction on T one can show that

$$\pi^*_{H(\mathfrak{A})}(\mathbf{u})\left(\pi^*_{H(\mathfrak{A})}(\mathbf{u})D(\mathbf{u})^{-1}\mathbf{x}\right)^2 = \sum_{t=1}^{T}\left(\frac{\mathbf{x}^{E_t}}{\mathbf{m}^{E_t}}\right)^2 - \sum_{t=2}^{T}\left(\frac{\mathbf{x}^{F_t}}{\mathbf{m}^{F_t}}\right)^2.$$

This proves (iii). □

It should be remarked that parts (ii) and (iii) of Lemma 6 become plausible, if one replaces $\log(1+h)$ by its first and second order Taylor expansion in (5.1). This will be exploited later.

PROOF OF PROPOSITION 3. Let

$$M = \max_{\substack{t=1,\dots,T\\ i\in I}}\left|\frac{n_i^{E_t} - m_i^{E_t}}{m_i^{E_t}}\right| = \max_{\substack{t=1,\dots,T\\ i\in I_t}}\left|\frac{n_i^{E_t} - m_i^{E_t}}{m_i^{E_t}}\right|,$$

where $I_t = \{i \in I | i_j = 1 \text{ for } j \in \{1,\dots,d\} \setminus E_t\}$. It follows from (5.1) that, for small M, $\|\hat{\mathbf{u}} - \mathbf{u}\|_\infty$ is bounded by const. M.

Here we use the boundedness of T and the fact that $F_t \subset E_t$ which allows to estimate $|(n_i^{F_t} - m_i^{F_t})/m_i^{F_t}| \le M$.

Let $r = \sum_{t=1}^T \prod_{j\in E_t} s_j = \sum_{t=1}^T |E_t|$. M is the maximum of r random variables. Furthermore let $\rho = \max_{t=1,\dots,T,\ i\in I}(1/m_i^{E_t})$. Since $H(E_t)$ is a subspace of $H(\mathfrak{A})$, we have $r = \sum_{t=1}^T \dim H(E_t) \le Tp$ and $1/m_i^{E_t} = \|\pi^*_{H(E_t)}(\mathbf{u})D(\mathbf{u})^{-1}\mathbf{e}^i\|_{\mathbf{u}} \le \kappa$. Therefore $\rho \le \kappa$ and r/p is bounded. It follows that $\rho \log r \to 0$.

We now prove that for all sequences C tending to $+\infty$,

$$P_{\mathbf{u}}\left(M > \sqrt{\rho \log r}\, C\right) \to 0, \tag{5.2}$$

which then proves that $\|\hat{\mathbf{u}} - \mathbf{u}\|_\infty$ tends to zero in probability at rate $\sqrt{\kappa \log p}$. Existence of the mle is guaranteed by (5.2) with probability tending to 1, since $\kappa \to 0$.

For a Poisson or binomial (univariate) random variable with expectation m the standard techniques of the theory of large deviations [exponentiation and estimation of the moment generating function, Bahadur (1971)] yield an estimation

$$P(|n - m| > b) \le 2\exp\left(-b^2/3m\right) \tag{5.3}$$

for all $0 \le b \le \varepsilon_0 m$, where ε_0 can be chosen independently of the distribution of n.

Now if $C \to \infty$, then also $\tilde{C} = C \wedge (\varepsilon_0/\sqrt{\rho \log r}) \to \infty$. We have

$$\begin{aligned} P_{\mathbf{u}}\left(M > \sqrt{\rho \log r}\, C\right) &\le P_{\mathbf{u}}\left(M > \sqrt{\rho \log r}\, \tilde{C}\right) \\ &= P_{\mathbf{u}}\left(\max_{\substack{t=1,\dots,T\\ i\in I_t}} \frac{1}{\sqrt{m_i^{E_t}}}\left|\frac{n_i^{E_t} - m_i^{E_t}}{\sqrt{m_i^{E_t}}}\right| > \sqrt{\rho \log r}\, \tilde{C}\right) \\ &\le \sum_{\substack{t=1,\dots,T\\ i\in I_t}} P_{\mathbf{u}}\left(|n_i^{E_t} - m_i^{E_t}| > \sqrt{m_i^{E_t} \log r}\, \tilde{C}\right). \end{aligned}$$

Observe that $\sqrt{m_i^{E_t} \log r}\,\tilde{C} \le m_i^{E_t}\sqrt{\rho \log r}\,\tilde{C} \le m_i^{E_t}\varepsilon_0$. Therefore we can apply (5.3) with $n = n_i^{E_t}$ and $b = \sqrt{m_i^{E_t} \log r}\,\tilde{C}$ to obtain

$$P_{\mathbf{u}}\big(M > \sqrt{\rho \log r}\, C\big) \le \sum_{\substack{t=1,\dots,T \\ i \in I_t}} 2\exp\left(-\frac{\tilde{C}^2 \log r}{3}\right)$$

$$= 2\exp\left(\left(1 - \frac{\tilde{C}^2}{3}\right)\log r\right),$$

which tends to zero. Therefore (5.2) holds. □

PROOF OF PROPOSITION 4. We only prove (a). The multivariate case can easily be reduced to be a univariate case as in the proof of Proposition 1.

If $f(E) = (\prod_{j \in \{1,\dots,d\}\setminus E} s_j)^{-1}$ denotes the redundancy factor introduced above, then for $\mathbf{z} \in H(E)$ it holds that

$$\sum_{i \in I} c_i z_i = f(E) \sum_{i \in I} c_i^E z_i.$$

Along with the estimation $|\log(1 + x) - (x - x^2/2)| \le C|x|^3$, valid for $|x| \le \frac{1}{2}$, this can be used in a straightforward calculation to show that

$$|(\mathbf{c}, \hat{\mathbf{u}} - \mathbf{u} - \mathbf{Q})|$$

$$\le C \sum_{t=1}^{T} f(E_t) \sum_{i \in I} |c_i^{E_t}| \left|\frac{n_i^{E_t} - m_i^{E_t}}{m_i^{E_t}}\right|^3 + C \sum_{t=2}^{T} f(F_t) \sum_{i \in I} |c_i^{F_t}| \left|\frac{n_i^{F_t} - m_i^{F_t}}{m_i^{F_t}}\right|^3 \tag{5.4}$$

$$\le M \cdot C \left\{ \sum_{t=1}^{T} f(E_t) \sum_{i \in I} |c_i^{E_t}| \left|\frac{n_i^{E_t} - m_i^{E_t}}{m_i^{E_t}}\right|^2 + \sum_{t=2}^{T} |c_i^{F_t}| \left|\frac{n_i^{F_t} - m_i^{F_t}}{m_i^{F_t}}\right|^2 \right\}.$$

[Hint: Use (5.1), Lemma 6(iii). W.l.o.g. $|M| \le \frac{1}{2}$.]

We prove that for $E \subset \{1,\dots,d\}$,

$$f(E) \sum_{i \in I} |c_i^E| \left|\frac{n_i^E - m_i^E}{m_i^E}\right|^2 = O_p\big(\sqrt{\kappa p}\big). \tag{5.5}$$

Since, by Proposition 3, $M = O_p(\sqrt{\kappa \log p})$ it then follows from (5.4) that

$$|(\mathbf{c}, \hat{\mathbf{u}} - \mathbf{u} - \mathbf{Q})| = O_p\big(\sqrt{\kappa \log p}\,\sqrt{\kappa p}\big) = O_p\big(\kappa\sqrt{p \log p}\big).$$

Since $\kappa\sqrt{p \log p} = \kappa p^{2/3}(\log p)^{1/3}(\log p/p)^{1/6} \to 0$, the first part of (a) is then proved.

To prove (5.5), the expectation ε of $f(E)\sum_{i \in I}|c_i^E|(|(n_i^E - m_i^E)/m_i^E|)^2$ is estimated. We have proved that $\pi_H^*(\mathbf{u})D(\mathbf{u})^{-1}\mathbf{x} = \mathbf{x}^E/\mathbf{m}^E$. Therefore

$$\varepsilon = f(E) \sum_{i \in I} |c_i^E| \cdot E\big(\pi_{H(E)}^*(\mathbf{u})D(\mathbf{u})^{-1}(\mathbf{n} - \mathbf{m}), \mathbf{e}^i\big)^2$$

$$\le f(E) \sum_{i \in I} |c_i^E| \, \|\pi_{H(E)}^*(\mathbf{u})D(\mathbf{u})^{-1}\mathbf{e}^i\|_{\mathbf{u}}^2$$

[compare with the estimation of $\mathrm{Var}(L^i)$ in the proof of Proposition 1]. An application of the Schwarz inequality yields

$$\varepsilon \leq \sqrt{\sum_{i \in I} m_i \left(\frac{c_i^E}{m_i^E} \right)^2} \sqrt{\sum_{i \in I} m_i \left\| \pi_{H(E)}^*(\mathbf{u}) D(\mathbf{u})^{-1} \mathbf{e}^i \right\|_{\mathbf{u}}^4}.$$

Now

$$\sum_{i \in I} m_i \left(\frac{c_i^E}{m_i^E} \right)^2 = \left\| \pi_{H(E)}^*(\mathbf{u}) D(\mathbf{u})^{-1} \mathbf{c} \right\|_{\mathbf{u}}^2 \leq 1,$$

since $H(E) \subset H(\mathfrak{A})$. Also

$$\sum_{i \in I} m_i \left\| \pi_{H(E)}^*(\mathbf{u}) D(\mathbf{u})^{-1} \mathbf{e}^i \right\|_{\mathbf{u}}^4 \leq \kappa \sum_{i \in I} m_i \left\| \pi_{H(E)}^*(\mathbf{u}) D(\mathbf{u})^{-1} \mathbf{e}^i \right\|_{\mathbf{u}}^2 = \kappa \dim H(E) \leq \kappa p.$$

Therefore (5.5) holds.

The second part of (a) follows from an application of the foregoing on $\tilde{\mathbf{u}}$, noting that $\tilde{\kappa}/\kappa \to 1$ in $P_{\mathbf{u}}$-probability, where $\tilde{\kappa} = \kappa(\tilde{\mathbf{u}})$. □

PROOF OF THE THEOREM. The convergence rate of $\|\hat{\mathbf{m}}/\mathbf{m} - 1\|_\infty$ is the same as that of $\|\hat{\mathbf{u}} - \mathbf{u}\|_\infty$. This was shown to be $\sqrt{\kappa \log p}$ in Proposition 3. Applying Proposition 1 with $\varepsilon = \|\hat{\mathbf{m}}/\mathbf{m} - 1\|_\infty$ yields

$$\begin{aligned} d_1((\mathbf{c}, \mathbf{Q}), (\mathbf{c}, \mathbf{Q}^*)) &= O_p\Big((\kappa \log p)^{1/4} \big(C_1 + C_2 (\kappa p)^{1/2} \big) \Big) \\ &= O_p\Big(C_1 (\kappa \log p)^{1/4} + C_2 \big(\kappa (\log p)^{1/3} p^{2/3} \big)^{3/4} \Big) \\ &= o_p(1). \end{aligned}$$

The error of replacing $\hat{\mathbf{u}} - \mathbf{u}$ by $\mathbf{Q}$ was estimated in Proposition 4.

The assertion about the multivariate distribution is proved analogously. □

6. Discussion. It is easy to construct examples in which the bootstrap works while the traditional approach does not. For instance let $\mathbf{n} = (n_{ij})$ be a sequence of $k \times k$ tables with $k \to \infty$ which follow the Poisson sampling model. Let H be the model of independence of the two factors [$H(\{1\}, \{2\})$ in our notation]. Assume that $m_{ij} = E(n_{ij}) = a$ for $i \leq k/2$ and $= 2a$ for $i > k/2$. A straightforward calculation [use Lemma 6(iv)] yields $\kappa = (ka)^{-1}(1 + 2/3 - 2/3k)$. Since $p = 2k - 1$ we may choose $a \to 0$ in such a way that condition (A) holds and consequently the bootstrap consistently estimates the distribution of standardized linear contrasts $\mathbf{c}$.

Now let $c_{ij} = 2(a/3k)^{1/2}$ for $i \leq k/2$ and $= -2(a/3k)^{1/2}$ for $i > k/2$. Then $\|\pi_H^*(\mathbf{u}) D(\mathbf{u})^{-1} \mathbf{c}\| = 1$. According to Proposition 4 $(\mathbf{c}, \hat{\mathbf{u}} - \mathbf{u})$ can be approximated by $(\mathbf{c}, \mathbf{Q})$. But for $b = a^{-1/2}/(2 \cdot 3^{1/2}) \to \infty$, $(\mathbf{c}, \mathbf{Q}) - b$ is asymptotically distributed as $N(0, 1)$.

The examples shows that the theory developed here covers large sparse contingency tables. It is somewhat artificial since a large (estimable) bias $b \to \infty$ only occurs in extremely large tables [in the example a necessary condition for (A) is $b^3/k \to 0$]. However, case studies performed by the author indicate that in large sparse tables occurring in practice a bias may be present and will be estimated by the bootstrap.

It is unknown to the author if nonnormal limit distributions occur under the condition of the theorem. This of course would give the results a larger practical importance. It is also unknown if condition (A) can be further weakened. As can be seen from the proof of the theorem, its special form derives from two components: the convergence rate of the mle and the bound on the Mallows-distance between the quadratic approximations of the mle's.

The difficulty in generalizing the results to indecomposable log-linear models lies in the fact that these models do not possess a closed-form mle. It is therefore not possible to prove Proposition 4 with the same techniques (Taylor's expansion and estimation of the remainder term). Instead methods like those developed by Haberman (1977a, b) would have to be used. If however the validity of the approximation of the mle by $\mathbf{Q}$ can be established, consistency of the bootstrap will follow from Proposition 1 which is valid for general (not only decomposable) log-linear models.

REFERENCES

ANDERSEN, A. H. (1974). Multidimensional contingency tables. *Scand. J. Statist.* **1** 115–127.

BAHADUR, R. R. (1971). *Some Limit Theorems in Statistics* SIAM, Philadelphia.

BERAN, R. (1984). Bootstrap methods in statistics. *Jahresber. Deutsch. Math.-Verein.* **86** 14–30.

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

BICKEL, P. J. and FREEDMAN, D. A. (1983). Bootstrapping regression models with many parameters. In *A Festschrift for Erich L. Lehmann* (P. J. Bickel, K. A. Doksum and J. L. Hodges, Jr., eds.) 28–48. Wadsworth, Belmont, Calif.

BISHOP, Y. M. M., FIENBERG, S. E. and HOLLAND, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, Mass.

DARROCH, J. N., LAURITZEN, S. L. and SPEED, T. P. (1980). Markov fields and log-linear interaction models for contingency tables. *Ann. Statist.* **8** 522–539.

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann Statist.* **7** 1–26.

EHM, W. (1986). On maximum likelihood estimation in high-dimensional log-linear models. I. The independent case. Preprint SFB 123, Univ. Heidelberg.

GOODMAN, L. A. (1970). The multivariate analysis of qualitative data: Interactions among multiple classifications. *J. Amer. Statist. Assoc.* **65** 226–256.

GOODMAN, L. A. (1971). Partitioning of chi-square, analysis of marginal contingency tables, and estimation of expected frequencies in multidimensional contingency tables. *J. Amer. Statist. Assoc.* **66** 339–344.

HABERMAN, S. (1970). The general log-linear model. Ph.D. dissertation, Dept. Statistics, Univ. Chicago.

HABERMAN, S. (1974). *The Analysis of Frequency Data*. Univ. Chicago Press, Chicago.

HABERMAN, S. (1977a). Maximum likelihood estimates in exponential response models. *Ann. Statist.* **5** 815–841.

HABERMAN, S. (1977b). Log-linear models and frequency tables with small expected cell counts. *Ann. Statist.* **5** 1148–1169.

HUBER, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821.

KOEHLER, K. J. (1986). Goodness-of-fit tests for log-linear models in sparse contingency tables. *J. Amer. Statist. Assoc.* **81** 483–493.

MALLOWS, C. L. (1972). A note on asymptotic joint normality. *Ann. Math. Statist.* **43** 508–515.

MORRIS, C. (1975). Central limit theorems for multinomial sums. *Ann. Statist.* **3** 165–188.

PORTNOY, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. Consistency. *Ann. Statist.* **12** 1298–1309.

PORTNOY, S. (1985). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. II. Normal approximation. *Ann. Statist.* **13** 1403–1417.

SHORACK, G. R. (1982). Bootstrapping robust regression. *Comm. Statist. A—Theory Methods* **11** 961–972.

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195.

GÖDECKE AG
BIOMETRICS DEPARTMENT
POSTFACH 569
MOOSWALDALLEE 1-9
D-7800 FREIBURG
WEST GERMANY