Goodness of fit tests and tests for independence of hierarchical log-linear models are studied for the special ease where samples are obtained from complex survey designs with fixed cell and marginal design effects. The asymptotic null distribution of the \( \chi^2 \) test is derived based on the results of Rao and Scott (1981, 1984). It is shown that, as a first order approximation, the usual \( \chi^2 \) test divided by the common cell and marginal design effect has a \( \chi^2 \) distribution.

### Case of constant design effects

For a given sampling design \( P(S) \), the design effect \( (def) \) of an estimate \( T \) is defined as the ratio of the variances of the estimate under the design \( P(S) \) and under simple random sampling:

\[
def(T) = \frac{\text{var}(T)}{\text{var}_{\text{srs}}(T)}
\]

where \( \text{var}_{\text{srs}}(\hat{p}_i) = \delta \text{var}_{\text{srs}}(\hat{p}_i) = \delta p_{0i}(1 - p_{0i})/n \) \( i = 1, 2, \ldots, k \) then the expected value of \( X_p^2 \) is

\[
E (X_p^2) = \sum_{i=1}^{k-1} \delta_i E (\chi^2_i)

= \sum_{i=1}^{k-1} \delta_i 1

= \text{trace}(P_0^{-1}V)

(\text{see Appendix 1})

Hence, \( X_p^2 \) has the same first moment as a \( \chi^2 \) random variable with \( k-1 \) degrees of freedom. Thus, when the design effects for cells are assumed to be constant, the Pearson statistic will have, as a first order approximation, the distribution of a \( \chi^2 \) random variable with

* It can be shown that except for the case \( k = 2 \), these assumptions do not imply that \( \text{var}(\hat{p}) \) is equal to \( \delta^{-1}P_0 \) (\( P_0 \) being the variance of \( \hat{p} \) under simple random sampling).
B. Log-Linear Model

In categorical data analysis, when the observations have more than one characteristic of interest, it is often the case that we would like to study how these characteristics interrelate. The study of these associations and interactions can be nicely formulated using a log-linear model.

Notation and background

Suppose that we have an r-dimensional contingency table with independent variables $x_1, x_2, \ldots, x_r$, each having respectively $v_1, v_2, \ldots, v_r$ categories. When $r = 3$, the indices $i, j, k$ can be used to denote a given cell in the table. For example, $\pi_{i,j,k}$ will denote the probability that an observation will be in the cell $i, j, k$. This notation can be generalized by using a single symbol, usually $\phi$, to denote the complete set of subscripts. Thus, $\pi_{\phi}$ will be the probability that an observation will be in an elementary cell $\phi$.

In this paper we will only consider hierarchical models as defined by Birch (1963). This means that the cell probabilities are permitted to be log-linearily related in such a way that a suitable set of marginals, usually called the minimal set of fitted marginals, is sufficient for the parameters. Tables of sums of non elementary cells in the sufficient configurations and will be called configurations and will be denoted by the letter $C$ (Bishop et al. (1975)). For example, in a three-way contingency table, the table of partial sums $\pi_{ij+} = \sum_k \pi_{i,j,k}$, obtained by summing over the third variable, will be denoted by $C_{12}$. As the third variable has been removed by summing, the subscripts of $C$ refer only to the remaining two variables. Configurations corresponding to the minimal set of fitted marginals, as defined above, will be called the sufficient configurations.

Bishop et al. (1975, page 68) outlined a method to derive sufficient configurations for comprehensive, unsaturated and hierarchical models. For such models, if the sufficient configurations are given, it is trivial to write down the log-likelihood function, log $m_\phi$. Also, it can be shown that the number of independent parameters in the model can be expressed in terms of the numbers of cells in the sufficient configurations.

Indeed, when only $C_0$ is the sufficient configuration of the model, it is clear that the number of independent variables in the model is equal to the number of cells in $C_0$. In other words, if $U_0$ is the set of all linearly independent $u$-terms whose subscripts are subsets of $\theta$ (which is, in this case, the set of all linearly independent parameters for the model) then the cardinality of $U_0$ is equal to the number of cells in $C_0$. This result implies that if the sufficient configurations are $C_{\theta_i}$, $i = 1, \ldots, k$ and the $U_{\theta_i}$'s sets are defined as above then $U = U_{\theta_1} \cup U_{\theta_2} \ldots \cup U_{\theta_k}$ will be the set of all linearly independent parameters and the cardinality of $U$ can be found using the inclusion-exclusion principle. For instance, the three dimensional contingency table with no three factor effect, log $m_{ijk} = u + u_{1(j)} + u_{2(k)} + u_{3(l)} + u_{12}(j) + u_{13}(k) + u_{23}(l)$ with $i = 1, \ldots, I$, $j = 1, \ldots, J$, $k = 1, \ldots, K$, has $C_{12}, C_{13}$ and $C_{23}$ as sufficient configurations. Let $U$ be the set of all linearly independent parameters then

$$\text{card}(U) = 1 + (I - 1) + (J - 1) + (K - 1) + (I - 1)(J - 1) + (I - 1)(K - 1) + (J - 1)(K - 1)$$

$$= IJ + JK + IK - I - J - K + 1$$

$$= \text{card}(U_{12}) + \text{card}(U_{13}) + \text{card}(U_{23}) - \text{card}(U_{12} \cap U_{13}) - \text{card}(U_{12} \cap U_{23}) - \text{card}(U_{13} \cap U_{23}) + \text{card}(U_{12} \cap U_{13} \cap U_{23}).$$

The formula for the number of independent variables will be simpler if the hierarchical log-linear model is decomposable. A hierarchical model with sufficient configurations $C_{\theta_i}$, $i = 1, \ldots, I$ is decomposable if and only if the class $\{\theta_i\}$ can be ordered in such a way that each $\theta_i$ is composed of one set of elements which are missing in all $\theta_s$ for $s > i$ and one set $\phi_i$ which is contained in some $\theta_s$ for some $\tau > i$ (Haberman (1974, chapter5), Sundberg (1975)). In other words, we have

$$\theta_i = \theta_i^* \cup \phi_i$$

with

$$\theta_i^* \cap \phi_s = \emptyset, \quad \theta_i^* \cap \bigcup_{j>i} \theta_j = \emptyset$$

$$\quad \forall s > i. \quad (\star)$$

Furthermore, it is a fact that if such an ordering is possible, a version may be found in which any prescribed set is the last one. For example,

(i) The three way contingency table with sufficient configuration $C_{12}, C_{13}$ and $C_{23}$ is not decomposable since the subscripts of any $C$ can not be decomposed into two disjoint subsets satisfying $(\star)$.

(ii) The seven dimensional hierarchical log-linear model with sufficient configurations $C_{12}, C_{13}, C_{23}, C_{136}, C_{138}$ and $C_{67}$ is decomposable. An ordering of the $\theta_i$ which has $\{5,7\}$ as the last set is

$$\{1,2,4\}, \{1,3,6\}, \{1,2,3\}, \{2,3,5\}, \{5,7\}$$

where the underlined elements do not belong to any set that follows. An ordering that has $\{1,3,6\}$ as the last set is

$$\{5,7\}, \{2,3,5\}, \{1,2,4\}, \{1,2,3\}, \{1,3,6\}$$

Usually, to obtain a particular ordering, it will be easier to start with the last set and work backwards.
Results under multinomial sampling

For completeness, let's first state some basic results about the log-linear model under multinomial sampling. The standard results for these models are given in Bishop et al. (1975) and Fienberg (1980). We will follow closely the notation in Rao and Scott's paper (1984). Let \( \pi = (\pi_1, \ldots, \pi_T)^T \) be a vector of cell proportions; \( \sum_{i=1}^{T} \pi_i = 1 \). We observe \( n = (n_1, \ldots, n_T)^T \) the counts in each cell from a random sample, so that \( n \) has a multinomial distribution (\( \sum n_i = N \)). Let \( p = n/N \) and define

\[
\mu = \log \pi
\]

The log-linear model assumes that for a parameter vector \( \theta = (\theta_1, \ldots, \theta_r)^T \), we have

\[
\mu(\theta) = u(\theta) 1 + X\theta,
\]

where \( X \) is a known \( T \times r \) matrix of full rank \( r \leq T-1 \) and \( X' 1 = 0 \), \( 1 \) is a \( T \)-vector of 1's. If \( r = T-1 \), we have a saturated model.

The maximum likelihood estimate for \( \theta \) is obtained by solving

\[
X^T(p - \hat{\pi}) = 0
\]

where \( \hat{\pi} = \pi(\hat{\theta}) \). Now asymptotically,

\[
n^{1/2}(\hat{\theta} - \theta) \to N(0, (X^T PX)^{-1})
\]

\[
n^{1/2}(\hat{\pi} - \pi) \to N(0, P(\pi X^T PX)^{-1} X^T P)
\]

in distribution.

Suppose now that the linear expression \( X \theta \) can be decomposed as \( X_1 \theta_1 + X_2 \theta_2 \) where \( X_1 \) and \( X_2 \) are full rank, \( X_1 \) is \( T \times s \), \( X_2 \) is \( T \times u \), \( \theta_1 \) is \( s \times 1 \) and \( \theta_2 \) is \( u \times 1 \) \( (s + u = r) \).

We consider the problem of testing

\( H_0 : \theta_2 = 0 \)

against the alternative

\( H_1 : \theta_2 \neq 0 \).

Let \( \hat{\theta}_1, \hat{\theta}_2, \hat{\pi} \), etc. be the maximum likelihood estimates under the full model \( H_1 \). Alternatively, let \( \tilde{\theta}_1, \tilde{\pi} \) denote the estimates under \( H_0 \). The likelihood ratio statistic for the above hypothesis is

\[
G^2 = 2n \sum \hat{p}_i \log(\hat{p}_i/\hat{\pi}_i) - 2n \sum \hat{p}_i \log(\hat{p}_i/\tilde{\pi}_i).
\]

Under \( H_0 \), this statistic has asymptotically a \( \chi^2 \) distribution with \( u \) degree of freedom. This statistic is also asymptotically equivalent to the Pearson statistic

\[
W_p = n(\pi - \hat{\pi})^T X^T P^{-1}(\pi - \hat{\pi})
\]

and the Wald statistic

\[
W_w = n\hat{\pi}^T X^T P X \hat{\theta}_2.
\]

Results for other sampling schemes

We still assume that the cell proportions, \( \pi \), satisfy

\[
\mu = \log \pi = u(\theta_1, \theta_2) 1 + X_1 \theta_1 + X_2 \theta_2
\]

but we now have

\[
n^{1/2}(p - \pi) \to N(0, V),
\]

where \( p \) is a survey estimate and \( V \) is the corresponding covariance matrix of \( p \). Rao and Scott (1984) showed that under general sampling designs, the test statistic \( X^T(\hat{\pi} - \pi) \) has asymptotically the distribution of the sum of weighted independent chi-squared variables with 1 degree of freedom,

\[
X^2 \sim \sum_{i=1}^{u} \delta_i W_i
\]

where the \( W_i \)'s are independent \( \chi^2 \) random variables and the \( \delta_i \)'s (all greater than 0) are the eigenvalues of

\[
(\hat{X}_2^T P \hat{X}_2)^{-1}(\hat{X}_2^T V \hat{X}_2)
\]

where

\[
\hat{X}_2 = (I - X_1 (X_1^T P X_1)^{-1} X_1^T P) X_2
\]

\[
P = D_\pi - \pi \pi^T, \quad D_\pi = \text{diag}(\pi).
\]

Rao and Scott also showed that under \( H_0 \),

\[
E(X^2) = E(G^2)
\]

\[
= E(G_1^2) - E(G_2^2)
\]

\[
= \sum_{i=1}^{u} \delta_i
\]

\[
= u \delta.
\]

\[
= \text{tr}(((X^T P X)^{-1}(X^T V X)) -
\]

\[
\text{tr}((X_1^T P X_1)^{-1}(X_1^T V X - 1))
\]

\( (G_1^2 \) and \( G_2^2 \) being the loglikelihood ratio under \( H_1 \) and \( H_0 \) respectively). Hence, as a first order approximation, \( \frac{X^2}{\delta} \) can be regarded as a \( \chi^2 \) with \( u \) degrees of freedom.

They also noted that when the models \( H_0 \) and \( H_1 \) admit explicit solutions for \( \hat{\pi} \) and \( \hat{\theta}_2 \), we have an alternative method of computing \( \delta_i \). Hierarchical log-linear models have closed form expressions for the maximum likelihood estimate for \( \pi \) only for decomposable models (Haberman (1974), Sundberg (1975)).

Let \( C_{\theta_1}, \ldots, C_{\theta_r} \) be the sufficient configurations for the model \( H_1 \) where the \( \theta_i \)'s are ordered according to the decomposability criterion, \( \phi_i = \theta_i \cap (\cup_{j<i} \theta_j) \) then the mle of \( \pi_\theta \) under \( H_1 \) is

\[
\hat{\pi}_\theta = \frac{1}{\prod_{i=1}^{I} \pi_{\phi_i}} \frac{1}{\prod_{i=1}^{I} \prod_{k \in Z - \cup_{j<i} \theta_j} \pi_{\phi_i}}
\]

where \( Z = \{1, 2, \ldots, r\} \) (Haberman (1974) and Sundberg (1975)).
In the above formula, \( \theta \) denotes an arbitrary cell, and \( \pi_{\theta_1}, \pi_{\theta_j} \) are marginal totals of \( \pi_{\theta} \) summed over indices not in \( \theta_1, \theta_j \) respectively. For example, in a five way table, for \( \theta = (2,4) \), \( \pi_\theta = \sum_{i_1, i_2, i_3} \pi_{i_1 i_2 i_3 i_4 i_5} \)

**Results when the cell and marginal design effects are equal**

Using a Taylor expansion, we have the following approximation for \( G_1^2 \),

\[
E(G_1^2) = \sum_{\theta} (1 - \pi_{\theta})d_{\theta} - \sum_{i} \sum_{\theta_i} (1 - \pi_{\theta_i})d_{\theta_i} + \sum_{j} \sum_{\theta_j} (1 - \pi_{\theta_j})d_{\theta_j}
\]

(Rao and Scott, 1984) where the \( d_{\theta_i}'s, d_{\theta_j}'s, d_{\phi_j}'s \) are the cell and marginal design effects. When the cell and design effects are all equal to \( \delta \), the above expression reduces to

\[
E(G_1^2) = \delta \left( \sum_{\theta} (1 - \pi_{\theta}) - \sum_{i} \sum_{\theta_i} (1 - \pi_{\theta_i}) + \sum_{j} \sum_{\theta_j} (1 - \pi_{\theta_j}) \right)
= \delta \left( T - 1 - \sum_{i=1}^{T} (# cells in \theta_i - 1) \right)
= \delta (T - \sum_{i=1}^{T} (# cells in \theta_i) + \sum_{j=1}^{T} (# cells in \theta_j))
= \delta (T - \# independent parameters in \theta)
\]

Similarly,

\[
E(G_2^2) = \delta (T - \# independent parameters in \theta)
\]

Thus,

\[
E(G^2) = u \delta,
\]

\[
E(G_2^2) = \delta (T - \# independent parameters in \theta)
\]

\[
E(G_1^2) = \delta \left( T - \sum_{i=1}^{T} (# cells in \theta_i) + \sum_{j=1}^{T} (# cells in \theta_j) \right)
= \delta \left( T - \# independent parameters in \theta \right)
\]

Hence, under \( H_0 \), \( X_2^2 = \frac{G_2^2}{\delta} \times \frac{\delta}{\lambda} \) has asymptotically a \( \chi^2 \) distribution with \( u \) degrees of freedom, where \( u \) is the difference of the number of independent parameters in the 2 models.

**References**


**Appendix 1**

Since \( P_0 = \text{diag}(\pi_0) - \pi_0 \pi_0' \), the inverse \( P_0^{-1} \) is

\[
P_0^{-1} = \left( \text{diag}(\pi_0) \right)^{-1} + \frac{1}{P_0} 11'
\]

Hence,

\[
P_0^{-1}V = \left( \text{diag}(\pi_0) \right)^{-1} V + \frac{1}{P_0} 11' V
\]

\[
\text{trace}(P_0^{-1}V) = \text{trace}(\text{diag}(\pi_0)^{-1} V) + \frac{1}{P_0} \text{var}(1 - p_01 - \ldots - p_{k-1})
\]

\[
= \sum_{i=1}^{k-1} \frac{v_{ii}}{p_0} + \frac{1}{P_0} \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} v_{ij}
\]

\[
= \sum_{i=1}^{k-1} \frac{v_{ii}}{p_0} + \frac{1}{P_0} \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} v_{ij}
\]

\[
= \sum_{i=1}^{k-1} \frac{v_{ii}}{p_0} + \frac{1}{P_0} \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} v_{ij}
\]

\[
= \sum_{i=1}^{k-1} \frac{v_{ii}}{p_0} + \frac{1}{P_0} \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} v_{ij}
\]

\[
= \sum_{i=1}^{k-1} \frac{v_{ii}}{p_0} + \frac{1}{P_0} \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} v_{ij}
\]

\[
= \sum_{i=1}^{k-1} \frac{v_{ii}}{p_0} + \frac{1}{P_0} \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} v_{ij}
\]

\[
= \sum_{i=1}^{k-1} \frac{v_{ii}}{p_0} + \frac{1}{P_0} \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} v_{ij}
\]