HILOGLINEAR

The Model Minimum Configuration

Consider an $I \times J \times K$ table. Let n_{ijk} be the observed frequency and m_{ijk} the expected frequency for cell (i, j, k). A simple way to construct a linear model in the natural logarithms of the expected cell frequencies is by analogy with analysis of variance (ANOVA) models:

$$L_{ijk} \equiv \log\left(m_{ijk}\right) = u + u_i^A + u_j^B + u_k^C + u_{ij}^{AB} + u_{ik}^{AC} + u_{jk}^{BC} + u_{ijk}^{ABC} \qquad \textbf{(1)}$$

where $1 \le i \le I, 1 \le j \le J$, and $1 \le k \le K$. In general, each of the seven subscripted *u-terms* sums to zero over each lettered subscript.

It can be shown (Bishop et al. (1975), p. 65) that, under the commonly encountered sampling plans, the log-likelihood is

$$\Phi + Nu + \sum_i n_{i++} u_i^A + \sum_j n_{+j+} u_j^B + \sum_k n_{++k} u_k^C + \sum_{i,j} n_{ij+} u_{ij}^{AB} + \sum_{i,k} n_{i+k} u_{ik}^{AC} + \sum_{j,k} n_{+jk} u_{jk}^{BC} + \sum_{i,j,k} n_{ijk} u_{ijk}^{ABC}$$

where Φ is independent of any parameters and *N* is total number of observations. Also, the *n*-terms adjacent to the unknown parameters are the sufficient statistics. The formulation of the above log likelihood is based on the saturated model **(1)**. When we consider unsaturated models, terms drop out and those that remain give the sufficient statistics. For instance, if we assume that there is no three-factor effect, that is, $u_{ijk}^{ABC} = 0$ for all *i*, *j*, and *k*, or more briefly $u_{ijk} = 0$, then

$$\log\left(m_{ijk}\right) = u + u_i^A + u_j^B + u_k^C + u_{ij}^{AB} + u_{ik}^{AC} + u_{jk}^{BC} \qquad \textbf{(2)}$$

and $n_{i++}, n_{+j+}, n_{++k}, n_{ij+}, n_{i+k}$ and n_{+jk} are the sufficient statistics for this reduced model. These statistics can be considered as *tables of sums configurations* and denoted by C with proper subscripts. For example, $\{n_{ij+}\}$ is the configuration C_{12} and $\{n_{++k}\}$ is the configuration C_3. Note that $\{n_{i++}\}, \{n_{+j+}\}$, and $\{n_{++k}\}$ can be obtained from $\{n_{ij+}\}, \{n_{i+k}\}$ and $\{n_{+jk}\}$. We then call the last three configurations C_{12}, C_{13} and C_{23} *minimal configurations or minimal statistics*.

Notation for Unlimited Number of Dimensions

To generalize results, we denote the complete set of subscripts by a single symbol θ. Thus, n_θ is the observed frequency in an elementary cell and w_θ is the cell weight. We add a subscript to θ to denote a reduced dimensionality so that $n_{\theta i}$ is the observed sum in a cell of $C_{\theta i}$. We use the second subscript, *i*, solely to distinguish between different configurations.

Iterative Proportional Fitting Procedure (IPFP)

We can obtain MLEs for the elementary cells under any hierarchical model by iterative fitting of the minimal sufficient configurations. To illustrate the algorithm, we consider model **(2)**. The MLEs must fit the configurations C_{12}, C_{13} and C_{23}. The basic IPFP chooses an initial table $m_{ijk}^{(0)}$ and then sequentially adjusts the preliminary estimates to fit C_{12}, C_{13}, and C_{23}. Fitting to C_{12} gives

$$\hat{m}_{ijk}^{(1)} = \hat{m}_{ijk}^{(0)} \frac{n_{ij+}}{\hat{m}_{ij+}^{(0)}}$$

Subsequent fitting to C_{13} gives

$$\hat{m}_{ijk}^{(2)} = \hat{m}_{ijk}^{(1)} \frac{n_{i+k}}{\hat{m}_{i+k}^{(1)}}$$

and similarly, after fitting C_{23} we have

$$\hat{m}_{ijk}^{(3)} = \hat{m}_{ijk}^{(2)} \frac{n_{+jk}}{\hat{m}_{+jk}^{(2)}}$$

We repeat this three-step cycle until convergence to the desired accuracy is attained. The extension of the above procedure to the general procedure for fitting *s* configurations is straightforward. Let the minimal configurations be $C_{\theta i}$ for $i = 1, 2, \ldots, s$, with cell entries $n_{\theta i}$, respectively. The procedure is as follows:

Initial Cell Estimates

To start the iterations, set

$$m_{\theta}^{(0)} = \begin{cases} 1 & \text{if } n_{\theta} > 0 \\ 0.5 & \text{if } n_{\theta} = 0 \\ 0 & \text{otherwise} \end{cases}$$

Intermediate Computations

After obtaining the initial cell estimates, the algorithm proceeds to fit each of these configurations in turn. After *r* cycles, the relations are

$$\hat{m}_{\theta}^{(sr+i)} = \hat{m}_{\theta}^{(sr+i-1)} \frac{n_{\theta_i}}{\hat{m}_{\theta_i}^{(sr+i-1)}} \quad \text{for } 1 \le i \le s; r \ge 0$$

Convergence Criteria

The computations stop either when a complete cycle, which consists of *s* steps, does not cause any cell to change by more than a preset amount ε; that is,

$$\left| \hat{m}_{\theta}^{(sr)} - \hat{m}_{\theta}^{(sr-s)} \right| < \varepsilon \quad \text{for all } \theta$$

or the number of cycles is larger than a preset integer *max*. Both ε and *max* can be specified. The default for ε is

$$\varepsilon = \max_{\theta}\left\{0.25; \frac{n_\theta w_\theta}{1000}\right\}$$

and the default for *max* is 20.

Goodness of Fit Tests

The Pearson chi-square statistic is

$$\chi^2 = \sum_{\theta} \frac{\left(n_\theta - \hat{m}_\theta\right)^2}{\hat{m}_\theta}$$

and the likelihood-ratio chi-square statistic is

$$L^2 = 2\sum_{\theta} n_\theta \ln\left(n_\theta / \hat{m}_\theta\right)$$

where the first summation is done over the cells with nonzero estimated cell frequencies while the second summation is done over cells with positive observed and estimated cell frequencies. The degrees of freedom for the above two statistics are computed as follows:

Adjusted Degrees of Freedom

Let T_c be the total number of the cells and P the number of parameters in the model. Also, let z_c be the number of cells such that $\hat{m}_\theta = 0$. The adjusted degrees of freedom is

$$\textbf{adjusted df} = T_c - P - z_c$$

Unadjusted Degrees of Freedom

unadjusted df $= T_c - P$

Parameter Estimates and Standard Errors

If a saturated model is fitted and neither $n_\theta + \delta$ nor w_θ is equal to zero for all cells, then the parameter estimates and their standard errors will be computed. Each estimate of the parameters in the saturated model can be expressed as a linear combination of the logarithms of the observed cell frequencies plus user-specified δ, where the coefficients used in the linear combination add to zero. We discuss the rule of obtaining the coefficients. Consider, in general case, a $J_1 \times J_2 \times \ldots \times J_M$ frequency table with defining variables $X_1, \ldots, X_M$. Let $u_{j_{s1}, \ldots, j_{sL}}^{X_{s1}, \ldots, X_{sL}}$ denote an *L*-term interaction involving $X_{s1}, \ldots, X_{sL}$ at level $j_{s1}, \ldots, j_{sL}$ respectively. Denote *A* as a vector that is constructed in the way that its nonzero components correspond to the variables in the parameter to be estimated and are set to the level of the variable. Let $C_{j_1, \ldots, j_M}$ be a *M*-dim vector with components equal to cell IDs. That is,

$$C_{j_1, \ldots, j_M} = (j_1, \ldots, j_M); \quad 1 \le j_1 \le J_1, \quad 1 \le i \le M$$

The coefficient $\beta_{j_1, \ldots, j_M}$ is determined through the comparison of the components of *A* and $C_{j_1, \ldots, j_M}$. Let *s* be the number of nonzero components of *A* that do not match (equal) the corresponding components of $C_{j_1, \ldots, j_M}$. Also, let matching occur at component $i_1, \ldots, i_k$. Then the coefficient for cell $(j_1, \ldots, j_M)$ is

$$\beta_{j_1, \ldots, j_M} = (-1)^s \left(J_{i_1} - 1\right) \times \ldots \times \left(J_{i_k} - 1\right)$$

The estimate $\hat{u}_{j_{s1},\ldots,j_{sL}}^{X_{s1},\ldots,X_{sL}}$ of $u_{j_{s1},\ldots,j_{sL}}^{X_{s1},\ldots,X_{sL}}$ is then

$$\hat{u}_{j_{s1},\ldots,j_{sL}}^{X_{s1},\ldots,X_{sL}} = \sum_{j_1,\ldots,j_M} \beta_{j_1,\ldots,j_M} \ln\left(n_{j_1,\ldots,j_M} + \delta\right) \quad \textbf{(3)}$$

The large-sample variance of the estimate **(3)** is

$$\sum_{j_1,\ldots,j_M} \beta_{j_1,\ldots,j_M}^2 \left[\ln\left(n_{j_1,\ldots,j_M} + \delta\right)\right]^{-1} \quad \textbf{(4)}$$

For a large sample, estimate **(3)** approximately follows a normal distribution with mean $u_{j_{s1},\ldots,j_{sL}}^{X_{s1},\ldots,X_{sL}}$ and variance **(4)** if the sampling model follows a Poisson, multinomial, or product-multinomial distribution. The confidence interval for the parameter can be computed based on the asymptotic normality.

Raw Residuals and Standardized Residuals

Two different residuals are computed:

Raw Residual

$$\text{raw residual} = n_\theta - \hat{m}_\theta$$

Standardized Residual

$$\text{standardized residual} = \left(n_\theta - \hat{m}_\theta\right) \Big/ \sqrt{\hat{m}_\theta}$$

where $\hat{m}_\theta$ must be greater than 0.

Partial Associations and Partial Chi-squares

Partial associations of effects can be requested when a saturated model is specified. Let $\chi^2(k)$ be the chi-square for the model that contains the effects up to and including the *k*-interaction terms. The test of the significance of the *k*th-order interaction can be based on

$$\chi^2(k-1) - \chi^2(k)$$

Degrees of freedom are obtained by subtracting the degrees of freedom for the corresponding models.

Model Selection Via Backward Elimination

The selection process starts with the model specified (either via DESIGN or MAXORDER subcommand). The partial chi-square is calculated for every term in the generating class. Any term with zero partial chi-square is deleted, then the effect with the largest observed significance level for the change in chi-square is deleted, provided the significance level is larger than 0.05, the default. With the removal of a highest-order term, a new model with new generating class is generated. The above process of removing a term is repeated for the new model and is continued until no remaining terms in the model can be deleted.

References

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. 1975. *Discrete multivariate analysis*: *Theory and practice*. Cambridge, Mass.: MIT Press.

Fienberg, S. E. 1977. *The analysis of cross-classified categorical data.* Cambridge, Mass.: The MIT Press.