1 Two-dimensional tables for nominal data:

Let X and Y be two nominal random variables with $I(A_i\,; i = 1, 2, .., I)$ and $J\,(B_j;\ j = 1, 2, .., J)$ categories respectively. Let m_{ij} denote the expected counts corresponding to the (ij)-th cell in a cross-classi¿ed table based on X and Y. That is, using X_{ij} to denote the number of sampled units with $X \in A_i$ and $Y \in B_j$, we have

$$m_{ij} = E(X_{ij})$$

where E is taken over the bivariate distribution of X and Y.

1.1 Three Sampling Schmes:

1. With no restriction on the total sample size, X_{ij} has an independent Poisson distribution. That is, in this case (two-dimensional table), the probability distribution of X_{ij} is given by

$$\Pr[X_{ij} = x_{ij}] = f(x_{ij}) = \frac{m_{ij}^{x_{ij}} e^{-m_{ij}}}{x_{ij}!};\ x_{ij} = 0, 1, 2, ...$$

Hence, the likelihood is given by

$$L = \prod_{i,j} \frac{m_{ij}^{x_{ij}} e^{-m_{ij}}}{x_{ij}!},$$

or log-likelihood is given by

$$l = \ln(L) = \sum_{ij} x_{ij} \ln(m_{ij}) - \sum_{ij} m_{ij} - \sum_{ij} \ln(x_{ij}!).$$

2. For a ¿xed sample size N, the joint distribution of X_{ij}'s will be a multinomial distribution given by

$$\Pr[X_{ij} = x_{ij}; i = 1, 2, ..I; j = 1, 2, .., J] = \frac{N!}{\prod_{i,j} x_{ij}!} \prod_{i,j} \left(\frac{m_{ij}}{N}\right)^{x_{ij}}.$$

Hence, the log-likelihood is given by

$$l = \ln(N!) + \sum_{ij} x_{ij} \ln(m_{ij}) - \sum_{ij} x_{ij} \ln(N) - \sum_{ij} \ln(x_{ij}!).$$

3. With fixed margins: First consider the case with row (X) margins fixed. That is, $m_{+j} = \sum_i m_{ij}$ are fixed (known) in advance prior to sampling. In this case, it can be shown that the joint distribution of X_{ij}'s will be a product of J independent multinomials, given by

$$\Pr[X_{ij} = x_{ij}; i = 1, 2, ..I; j = 1, 2, .., J] = \prod_{j,} [\frac{m_{+j}!}{\prod_{i,} x_{ij}!} \prod_{i,} \left(\frac{m_{ij}}{m_{+j}} \right)^{x_{ij}}.$$

Hence, the log-likelihood is given by

$$l = \sum_j \ln(m_{+j}!) + \sum_{ij} x_{ij} \ln(m_{ij}) - \sum_j x_{+j} \ln(m_{+j}) - \sum_{ij} \ln(x_{ij}!).$$

Similarly, for the case with column (Y) margins fixed (that is when $m_{i+} = \sum_j m_{ij}$ are known), the log-likelihood is given by

$$l = \sum_i \ln(m_{i+}!) + \sum_{ij} x_{ij} \ln(m_{ij}) - \sum_i x_{i+} \ln(m_{i+}) - \sum_{ij} \ln(x_{ij}!).$$

Note that in all above cases, the kernel (the term containing both x_{ij} and m_{ij}) is

$$l^* = \sum_{ij} x_{ij} \ln(m_{ij}).$$

Hence, maximizing this kernel with respect to m_{ij} with no constraints on m_{ij} for the first sampling scheme; with constraints $\sum_{ij} m_{ij} = N$ under the second sampling scheme; and with constraints $\sum_i m_{ij} = m_{+j}$ for fixed row margin case, and $\sum_j m_{ij} = m_{i+}$ for fixed column margin case will provide maximum likelihood estimates for m_{ij}'s under respective cases.

Also, by noting that Poisson and multinomial belong to the class of exponential PDF's, the quantities $\ln(m_{ij})$ are called canonical parameters under above sampling schemes. Since the data collected from any of the above sampling schemes can be treated under one basic model Birch introduced log-linear parameterization which is essentially a reparametrization of parameters under different models in terms of the canonical parameters $\ln(m_{ij})$ under the basic model. In this case the reparmetrization is given by

$$\ln(m_{ij}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}, \tag{1}$$

where the λ's satisfy the linear constraints

$$\sum_i \lambda_{ij}^{XY} = 0 = \sum_j \lambda_{ij}^{XY}; \sum_i \lambda_i^X = 0 = \sum_j \lambda_j^Y.$$

That is , there (I-1)(J-1) independent λ_{ij}^{XY}'s, (I-1) independent λ_i^X's , and (J-1) independent λ_j^Y's parameters.

Hypothesis of No Interaction:

$$H_0 : \lambda_{ij}^{XY} = 0; i = 1, 2, ..I, j = 1, 2, .., J.$$

Under the Poisson sampling scheme the above hypothesis is known the hypothesis of the multiplicative Poisson model. This can be verified by noting that under H_0,

$$m_{ij} = e^{\mu} e^{\lambda_i^X} e^{\lambda_j^Y}.$$

For the mulinomial sampling (with overall sample size N,fixed) the above hypothesis represents the hypothesis of independence between X and Y. Where as under the sampling schemes with fixed row or column margins the above hypothesis is equivalent to the hypothesis of homogeneity.

Sufficient Statistics:

Using the above log-linear parametrization, we can write the kernel of the log-likelihood under all sampling schemes as

$$l^* = \sum_{ij} x_{ij} \ln(m_{ij}) = \mu \sum_{ij} x_{ij} + \sum_{i} x_{i+} \lambda_i^X + \sum_{j} x_{+j} \lambda_j^Y + \sum_{ij} x_{ij} \lambda_{ij}^{XY} \qquad (2)$$

Since the underlying probability models under the above sampling schemes belong to the family of exponential PDF's, the sufficient statistics in these cases are the x-terms adjacent to the unknown parameters, $\lambda-$terms. Thus, for the saturated model (1) for a two-dimensional table $\{x_{ij}; i = 1, 2, .., I, j = 1, 2, .., J\}$ is the minimal sufficient statistic. Under $H_0 : \lambda_{ij}^{XY} = 0; i = 1, 2, .., I, j = 1, 2, .., J$, the kernel of log-likelihood for the model (1) reduces to

$$l^* = \sum_{ij} x_{ij} \ln(m_{ij}) = \mu \sum_{ij} x_{ij} + \sum_{i} x_{i+} \lambda_i^X + \sum_{j} x_{+j} \lambda_j^Y.$$

Therefore, $[\{x_{i+}; i = 1, 2, ..., I\}, \{x_{+j}; j = 1, 2, ..., J\}]$ is mimimal sufficient statistic.

Following Birch's results (stated later for a more general model) the maximum likelihood estimates for unknown parametes under H_0 are obtained by solving the following equations for $\hat{m}_{ij}$:

$$\begin{aligned} \hat{m}_{i+} &= x_{i+} \\ \hat{m}_{+j} &= x_{+j}. \end{aligned}$$

That is, the maximum likelihood estimates under H_0 are given by

$$\hat{m}_{ij0} = \frac{x_{i+}x_{+j}}{x_{++}}.$$

Note that for the saturated model, since $\{x_{ij}\}$ is the minimal sufficient statistics, the m.l.e are given by

$$\hat{m}_{ij} = x_{ij}.$$

Hence, the G^2 and X^2 statistics for testing H_0 are respectively given by

$$G^2 = -2\sum_{ij} x_{ij}\ln\left(\frac{\hat{m}_{ij0}}{\hat{m}_{ij}}\right) = 2\left[\sum_{ij} x_{ij}\ln(x_{ij}) - \sum_{i} x_{i+}\ln(x_{i+}) - \sum_{j} x_{+j}\ln(x_{+j}) + N\ln(N)\right]$$

and

$$X^2 = \sum_{ij}\frac{\left(x_{ij}-\hat{m}_{ij0}\right)^2}{\hat{m}_{ij0}}.$$

When H_0 is true, both statistics are distributed as χ^2 with degrees of freedom equal to $(I-1)(J-1)$. It can be shown that (see Lemma 14.9_1; BFH, page 514)

$$G^2 = X^2 + O_p(N^{-1/2}).$$

Relationships between λ-terms: Consider two λ-terms, one with r subscripts and the other with s subscripts, where $\dot{r} > s$. Then these two terms are relatives if the r subscripts contain among them all the s subscripts, and the term with r subscripts is called a higher order relative term. For example, in a two-dimensional table model λ_{ij}^{XY} is higher order relative of both λ_i^X and λ_j^Y terms.

The hierarchy principle:

The family of hierarchical models is defined as the family of log-linear models such that if any $\lambda-$term is set equal to zero, all its higer-order relatives must also be set equal to zero. Conversely, if any $\lambda-$term is not zero, its lower-order relatives must be present in the log-linear model.

2 Three dimensional Tables:

Consider three categorical variables X, Y, Z , respectively, having I, J and K categories. With sample size N we have a three-way table of counts by cross-clasifying X, Y and Z, and denote a typical count n_{ijk} where $i = 1, 2, .., I$,

$j = 1, 2, ..., J$,and $k = 1, 2, ..., K$. Similarly denote the cell probability (the probability that an observation falls in the given cell) by p_{ijk} and the expected cell count as m_{ijk}. The saturated model in this is given by

$$\ln(m_{ijk}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}$$

with suitable constraints as in the case of two-dimensional table. The above model consists of a list of terms, called generators, corresponding to the maximal interaction term XYZ in the model. Following the heirachy principle, this term uniquely de¿nes the above model. Hence, this maximal interaction term is called the generator of the model. Now, consider a simple model

$$\ln(m_{ijk}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}.$$

Note that in the above model maximal interaction terms are XY, XZ and YZ. Hence, in this case we call XY, XZ and XZ as generators of the model.

Mutually Independent Model: If the model containing only main e¾ects (i.e., when all interactions are absent) is the best ¿tted model then the variables X, Y and Z are said to be mutually independently distributed. This can be seen as follows. Consider the main e¾ects model, that is the model with generators X, Y and Z,

$$\ln(m_{ijk}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z.$$

Under this model, we can note that

$$p_{ijk} = (p_{i++})(p_{+j+})(p_{++k}),$$

where $p_{i++} = \sum_{jk} p_{ijk}$, $p_{+j+} = \sum_{ik} p_{ijk}$ and $p_{++k} = \sum_{ij} p_{ijk}$. Also, from above equation it follows

$$\sum_k p_{ijk} = p_{ij+} = p_{i++}p_{+j+}$$

which implies X and Y are independent. Similarly independence between X and Z, and between Y and Z follows.

3 Multidimensional Tables:

Extending the results of the previous sections to multidimensional tables is quite straightforward except for notational diÁculty.

NOTATIONS:

Let

d =dimension of a table

Δ =Set of d categorical variables.

$I_j = \#$ of categories associated with the j-th variable

$\theta = \{i_1 i_2 ... i_d\}$ the complete set of subscripts, where $i_j = 1, 2, ..., I_j$.

$\eta = \#$ of subsets $\theta_k \subseteq \theta$

λ_i^a =general interaction term with set of variables defined by $a \subseteq \Delta$. It is understood here that λ_i^a depends on i only through i_a where i_a is a sub $d-$tuple of i.

c =# of sufficient configurations $(\theta_k \subseteq \theta; k = 1, 2, ...\eta)$

C_{θ_i} =configuration corresponding to $\theta_i, i = 1, 2, ..., \eta$

x_θ =observed count in an elementary cell

x_{θ_i} = observed count in a cell defined by the configuration C_{θ_i}

m_θ =expected count in an elementary cell

$\ln(m_\theta) = \sum_{a \subseteq \Delta} \lambda_i^a$, the full (saturated model)

m_{θ_i} =expected count in a cell defined by the configuration C_{θ_i}

$\hat{m}_\theta$ = The m.l.e of m_θ

$\hat{m}_{\theta_i}$ = The m.l.e of m_{θ_i}

$N = \sum_\theta x_\theta$ =Sample size.

3.1 Steps for generating sufficient configurations and sufficient statistics for hierarchical models:

(i). Select $\lambda-$terms of highest order interaction, say t, in the model (t$\leq d$)

(ii). If all possible interactions of order t (there are $\binom{d}{t}$ interaction terms of order t) are included in the model, stop selection with configurations corresponding to these interactions giving $\{C_l\}$.

(iii) Otherwise, continue by examining terms of order $(t-1)$ and select those that are not lower order relatives of terms of order t in the model.

(iv). Continue this process for λ-terms of every order and select at each step only those terms not included in higher order terms.