二、三维列联表确切概率检验的有关算法*

李 建 立
(上海第二医科大学生物统计教研室)

AN ALGORITHM IN THE EXACT PROBABILITY TEST IN TWO-AND THREE-DIMENSIONAL CONTINGENCY TABLES

Li Jian-li
(*Section of Biostatistics Shanghai Second Medical University*)

Abstract

In this paper an algorithm for finding out all possible two-and three-dimensional contingency tables with the fixed marginals is given. In this algorithm, a network consisting of nodes and arcs is constructed and several paths are set up. Corresponding to each node and each path a unique nonnegative value is defined and the conditions for these values are given. From these values all possible contingency tables can be found. Furthermore, the formulas of calculating P values in exact probability tests are given.

对于一个列联表当样本较小时，就会产生用 χ^2 统计量或 G^2 统计量来作关于交互作用的显著性检验不够精确的问题，于是就需作确切概率检验。本文就二、三维列联表在某些条件下的确切概率计算的有关算法加以讨论。

§1. 列联表的样本分布

若对所抽的样本大小没有任何约束，且每个格子的观察频数 $\{X_\theta\}$ 可看作服从独立的 Poisson 分布。各自的均值为 $\{m_\theta\}$，其中 θ 表示下标集：$\{i_1, i_2, \cdots, i_n\}$。对于一个 n 维列联表 $\{X_\theta\}$，其密度函数(联合分布密度函数)为

$$f(\{X_\theta\}) = \prod_\theta \frac{m_\theta^{X_\theta} e^{-m_\theta}}{X_\theta !}. \tag{1}$$

* 1988 年 11 月 11 日收到。

有时往往要考虑样本大小固定的独立 Poisson 抽样，假设 $\sum_\theta X_\theta = N$. 可知 $\{X_\theta\}$ 的分布为多项分布，其密度函数为

$$f(\{X_\theta\}) = \frac{N!}{\prod_\theta X_\theta!} \prod_\theta \left(\frac{m_\theta}{N}\right)^{X_\theta}, \tag{2}$$

其中 m_θ 为 X_θ 的期望值，$m_\theta/N = p_\theta \quad \left(\sum_\theta p_\theta = 1\right)$.

在实验设计中，会确定各组的总数，亦即固定组态(边际和)大小 $c_{\theta_i}(\theta_i \subseteq \theta)$，$c_{\theta_i}$ 即 $\left\{\sum_{\theta-\theta_i} X_\theta\right\}$ (其中 $\theta-\theta_i$ 表示下标集 θ 与下标集 θ_i 之差. $\sum_{\theta-\theta_i}$ 表示对 $\theta-\theta_i$ 中所有下标求和)，其结果得到的分布是多个多项分布之积. 推导如下：

因为 c_{θ_i} 的边际分布为

$$f\left(\left\{\sum_{\theta-\theta_i} X_\theta\right\}\right) = \frac{N!}{\prod_{\theta_i}\left(\sum_{\theta-\theta_i} X_\theta\right)!} \prod_{\theta_i}\left(\frac{\sum_{\theta-\theta_i} m_\theta}{N}\right)^{\sum_{\theta-\theta_i} X_\theta},$$

所以条件分布为

$$f\left(\{X_\theta\} \,\middle|\, \left\{\sum_{\theta-\theta_i} m_\theta = \sum_{\theta-\theta_i} X_\theta\right\}\right) = f(\{X_\theta\}) \Big/ f\left(\left\{\sum_{\theta-\theta_i} X_\theta\right\}\right)$$

$$= \prod_{\theta_i}\left[\frac{\left(\sum_{\theta-\theta_i} X_\theta\right)!}{\prod_{\theta-\theta_i} X_\theta!} \prod_{\theta-\theta_i}\left(\frac{m_\theta}{\sum_{\theta-\theta_i} X_\theta}\right)^{X_\theta}\right]. \tag{3}$$

§2. 二、三维表的确切概率计算

二维列联表 $\{X_{ij}\}$ 有以下结果[1]：

结果 1. 若原假设 $H_0: u_{12}=0$ 成立，即二变量相互独立，且有条件：边际和 $X_{i.}$ 与 $X_{.j}$ 固定，则分布密度函数为

$$f^*(\{X_{ij}\}) = \frac{\prod_i X_{i.}! \prod_j X_{.j}!}{\prod_{ij} X_{ij}! N!}, \tag{4}$$

其中 $X_{i.} = \sum_j X_{ij}, X_{.j} = \sum_i X_{ij},\ N = X_{..} = \sum_{ij} X_{ij}$.

为考察观察到的列联表 $\{X^*_{ij}\}$ 的条件显著性，可计算下述概率值：

$$P = S(\{X^*_{ij}\}) = \sum f^*(\{X'_{ij}\}), \tag{5}$$

其中求和是对所有 $\{X'_{ij}\}$ 进行的，而 $f^*(\{X'_{ij}\}) \leqslant f^*(\{X^*_{ij}\})$.

根据(4)式，因为 $X_{i\cdot}$ 与 $X_{\cdot j}$ 固定，所以上述选择 $\{X'_{ij}\}$ 的条件可改写为

$$\frac{1}{\prod\limits_{ij} X'_{ij}!} \leqslant \frac{1}{\prod\limits_{ij} X^*_{ij}!}, \tag{6}$$

即

$$\prod_{ij} \frac{X'_{ij}!}{X^*_{ij}!} \geqslant 1. \tag{7}$$

对于事先给定检验水平 α，作双侧检验，若 $P \leqslant \alpha$，则拒绝原假设。

三维列联表 $\{X_{ijk}\}$ 的情况较复杂，本文只就二种情况作讨论，有以下结果：

结果 2. 若原假设 $H_0: u_{12}=u_{13}=u_{23}=u_{123}=0$ 成立，即三变量相互独立，且有条件：边际和 $X_{i\cdot\cdot}$、$X_{\cdot j\cdot}$ 及 $X_{\cdot\cdot k}$ 固定，则分布密度函数为

$$f^*(\{X_{ijk}\}) = \frac{\prod\limits_i X_{i\cdot\cdot}! \prod\limits_j X_{\cdot j\cdot}! \prod\limits_k X_{\cdot\cdot k}!}{\prod\limits_{ijk} X_{ijk}!(N!)^2}, \tag{8}$$

其中 $N = X_{\cdot\cdot\cdot} = \sum\limits_{ijk} X_{ijk}, X_{i\cdot\cdot} = \sum\limits_{jk} X_{ijk}, X_{\cdot j\cdot} = \sum\limits_{ik} X_{ijk},\ X_{\cdot\cdot k} = \sum\limits_{ij} X_{ijk}$.

推导如下：

根据边际和固定条件，若原假设成立，即三变量相互独立，可推得

$$\begin{aligned} &f(\{X_{ijk}\} \mid \{m_{i\cdot\cdot} = X_{i\cdot\cdot}, m_{\cdot j\cdot} = X_{\cdot j\cdot}, m_{\cdot\cdot k} = X_{\cdot\cdot k}\}) \\ &= f(\{X_{ijk}\} \mid \{m_{i\cdot\cdot} = X_{i\cdot\cdot}\}) / f(\{X_{\cdot j\cdot}\}) / f(\{X_{\cdot\cdot k}\}) \\ &= \frac{\prod\limits_i X_{i\cdot\cdot}! \prod\limits_{ijk} (m_{ijk})^{X_{ijk}} \prod\limits_j X_{\cdot j\cdot}! N^N \prod\limits_k X_{\cdot\cdot k}! N^N}{\prod\limits_{ijk} X_{ijk}! \prod\limits_i (X_{i\cdot\cdot})^{X_{i\cdot\cdot}} N! \prod\limits_j (m_{\cdot j\cdot})^{X_{\cdot j\cdot}} N! \prod\limits_k (m_{\cdot\cdot k})^{X_{\cdot\cdot k}}}. \end{aligned} \tag{9}$$

当 H_0 成立时，应有 $m_{ijk} = \dfrac{X_{i\cdot\cdot} X_{\cdot j\cdot} X_{\cdot\cdot k}}{N^2}$，即可由(9)推得

$$\begin{aligned} &f^*(\{X_{ijk}\}) \\ &= \frac{\prod\limits_i X_{i\cdot\cdot}! \prod\limits_j X_{\cdot j\cdot}! \prod\limits_k X_{\cdot\cdot k}! \prod\limits_i (X_{i\cdot\cdot})^{X_{i\cdot\cdot}} \prod\limits_j (X_{\cdot j\cdot})^{X_{\cdot j\cdot}} \prod\limits_k (X_{\cdot\cdot k})^{X_{\cdot\cdot k}} N^{2N}}{\prod\limits_{ijk} X_{ijk}!(N!)^2 N^{2N} \prod\limits_i (X_{i\cdot\cdot})^{X_{i\cdot\cdot}} \prod\limits_j (X_{\cdot j\cdot})^{X_{\cdot j\cdot}} \prod\limits_k (X_{\cdot\cdot k})^{X_{\cdot\cdot k}}} \\ &= \frac{\prod\limits_i X_{i\cdot\cdot}! \prod\limits_j X_{\cdot j\cdot}! \prod\limits_k X_{\cdot\cdot k}!}{\prod\limits_{ijk} X_{ijk}!(N!)^2}. \end{aligned} \tag{10}$$

对于一个观察到的列联表 $\{X^*_{ijk}\}$，在原假设(三变量相互独立)成立及边际和 $X_{i\cdot\cdot}$，$X_{\cdot j\cdot}$，$X_{\cdot\cdot k}$ 固定的条件下的条件概率值记为 $f^*(\{X^*_{ijk}\})$.

为考察观察到的列联表 $\{X^*_{ijk}\}$ 的条件显著性，可计算下述概率值：

$$P=S(\{X^*_{ijk}\})=\sum f^*(\{X'_{ijk}\}),\tag{11}$$

其中求和是对所有 $\{X'_{ijk}\}$ 进行的，而 $f^*(\{X'_{ijk}\})\leqslant f^*(\{X^*_{ijk}\})$.

根据(11)式，因为 $X_{i..},X_{.j.},X_{..k}$ 固定，所以上述选择 $\{X'_{ijk}\}$ 的条件可改写为

$$\prod_{ijk}\frac{X'_{ijk}!}{X^*_{ijk}!}\geqslant 1.\tag{12}$$

对于事先给定检验水平 α，作双侧检验，若 $P\leqslant\alpha$，则拒绝原假设.

结果 3. 若原假设 $H_0:u_{12}=u_{13}=u_{123}=0$ 成立，即二变量（变量 1 与变量 2 及变量 1 与变量 3）完全独立，且有条件：边际和 $X_{i..}$ 与 $X_{.jk}$ 固定，则分布密度函数为

$$f^*(\{X_{ijk}\})=\frac{\prod\limits_{i}X_{i..}!\prod\limits_{jk}X_{.jk}!}{\prod\limits_{ijk}X_{ijk}!N!}.\tag{13}$$

推导如下：

根据边际和固定条件，若原假设成立，则可推得

$$\begin{aligned}&f(\{X_{ijk}\}\mid\{m_{i..}=X_{i..},m_{.jk}=X_{.jk}\})\\&\quad=f(\{X_{ijk}\}\mid\{m_{i..}=X_{i..}\})/f(\{X_{.jk}\})\\&\quad=\frac{\prod\limits_{i}X_{i..}!\prod\limits_{ijk}(m_{ijk})^{X_{ijk}}\prod\limits_{jk}X_{.jk}!N^N}{\prod\limits_{ijk}X_{ijk}!\prod\limits_{i}(X_{i..})^{X_{i..}}\prod\limits_{jk}(m_{.jk})^{X_{.jk}}N!}.\end{aligned}\tag{14}$$

当 H_0 成立时，应有 $m_{ijk}=\dfrac{X_{i..}X_{.jk}}{N}$，即可由(14)推得

$$\begin{aligned}f^*(\{X_{ijk}\})&=\frac{\prod\limits_{i}X_{i..}!\prod\limits_{i}(X_{i..})^{X_{i..}}\prod\limits_{jk}(X_{.jk})^{X_{.jk}}\prod\limits_{jk}X_{.jk}!N^N}{\prod\limits_{ijk}X_{ijk}!N^NN!\prod\limits_{jk}(X_{.jk})^{X_{.jk}}\prod\limits_{i}(X_{i..})^{X_{i..}}}\\&=\frac{\prod\limits_{i}X_{i..}!\prod\limits_{jk}X_{.jk}!}{\prod\limits_{ijk}X_{ijk}!N!}.\end{aligned}\tag{15}$$

同样，若要考察观察到的 $\{X^*_{ijk}\}$ 的条件显著性，只需计算概率值

$$P=S(\{X^*_{ijk}\})=\sum_{f^*(\{X'_{ijk}\})\leqslant f^*(\{X^*_{ijk}\})}f^*(\{X^*_{ijk}\}).$$

易见，选择 $\{X'_{ijk}\}$ 之条件为

$$\prod_{ijk}\frac{X'_{ijk}!}{X^*_{ijk}!}\geqslant 1.\tag{16}$$

§3. 产生各种可能的列联表的算法

在列联表某些边际和固定的情况下，可用网络来求得各种可能的列联表，此网络由节点和连线组成。

对于二维列联表 $\{X_{ij}\}(1\leqslant i\leqslant I, 1\leqslant j\leqslant J)$，每一节点记为 $ND_{ij}^{[\theta_{ij}]}(1\leqslant i\leqslant I-1,\ 0\leqslant j\leqslant J)$，其中 $[\theta_{ij}]=n_{ij}n_{i,j-1},\cdots n_{i0}\cdots n_{10}$ 为节点序列号，且可有以下分解：$[\theta_{ij}]=n_{ij}[\theta_{i,j-1}]$。对于节点 $ND_{ij}^{[\theta_{ij}]}$ 和 $ND_{i,j-1}^{[\theta_{i,j-1}]^*}$，若有 $[\theta_{i,j-1}]^*=[\theta_{i,j-1}]$，则这两节点间可有连线相联，这些连线构成了通路。对应于每一个节点和每条通路有唯一确定的一个非负整数值，记为 $V_{ij}^{(\alpha)}$，其中 α 表示通路号，且令 $V_{i0}^{(\alpha)}=0$。

定义 1. 二维列联表 $\{X_{ij}\}$ 中第 (i,j) 格第 α 种可能的频数

$$X_{ij}^{(\alpha)}=V_{ij}^{(\alpha)}-V_{i,j-1}^{(\alpha)}.$$

引理 1. 二维列联表 $\{X_{ij}\}(1\leqslant i\leqslant 2,\ 1\leqslant j\leqslant J)$ 中，若 $X_{i\cdot}$ 与 $X_{\cdot j}$ 固定，则 $V_{ij}^{(\alpha)}$ 满足以下条件：

$$\max\left(V_{1,j-1}^{(\alpha)},\ \sum_{l=1}^{j}X_{\cdot l}-X_{2\cdot}\right)\leqslant V_{1j}^{(\alpha)}\leqslant\min\left(V_{1,j-1}^{(\alpha)}+X_{\cdot j},X_{1\cdot}\right).\tag{17}$$

证. 在 $\{X_{ij}\}$ 中，若 $X_{1l}^{(\alpha)}(1\leqslant l\leqslant j-1)$ 已确定，则显然有

$$X_{1j}^{(\alpha)}\leqslant X_{1\cdot}-\sum_{l=1}^{j-1}X_{1l}^{(\alpha)}.$$

根据定义 1 有

$$V_{1j}^{(\alpha)}-V_{1,j-1}^{(\alpha)}\leqslant X_{1\cdot}-\sum_{l=1}^{j-1}\left(V_{1l}^{(\alpha)}-V_{1,l-1}^{(\alpha)}\right),$$

$$V_{1j}^{(\alpha)}\leqslant X_{1\cdot}-\sum_{l=1}^{j-1}\left(V_{1l}^{(\alpha)}-V_{1,l-1}^{(\alpha)}\right)+V_{1,j-1}^{(\alpha)},$$

所以

$$V_{1j}^{(\alpha)}\leqslant X_{1\cdot}.\tag{18}$$

又因为 $X_{1j}^{(\alpha)}\leqslant X_{\cdot j}$，所以 $V_{1j}^{(\alpha)}-V_{1,j-1}^{(\alpha)}\leqslant X_{\cdot j}$，即有

$$V_{1j}^{(\alpha)}\leqslant X_{\cdot j}+V_{1,j-1}^{(\alpha)}.\tag{19}$$

根据(18)和(19)有

$$V_{1j}^{(\alpha)}\leqslant\min\left(V_{1,j-1}^{(\alpha)}+X_{\cdot j},X_{1\cdot}\right).\tag{20}$$

另一方面，由定义 1 可知

$$V_{1j}^{(\alpha)}-V_{1,j-1}^{(\alpha)}=X_{1j}^{(\alpha)}\geqslant 0,$$

即

$$V_{1j}^{(\alpha)}\geqslant V_{1,j-1}^{(\alpha)}.\tag{21}$$

又因为 $\sum_{l=1}^{j}(X_{\cdot l}-X_{1l})\leqslant X_{2\cdot}$，即 $\sum_{l=1}^{j}X_{\cdot l}-\sum_{l=1}^{j}\left(V_{1l}^{(\alpha)}-V_{1,l-1}^{(\alpha)}\right)\leqslant X_{2\cdot}$，所以

$$V_{1j}^{(\alpha)} \geqslant \sum_{l=1}^{j} X_{\cdot l} - X_{2\cdot\cdot} \tag{22}$$

根据(21)和(22)有

$$V_{1j}^{(\alpha)} \geqslant \max\left(V_{1,j-1}^{(\alpha)}, \sum_{l=1}^{j} X_{\cdot l} - X_{2\cdot}\right). \tag{23}$$

由(20)和(23)可得结论。

定理 1. 二维列联表 $\{X_{ij}\}$ $(1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J)$ 中，若 $X_{i\cdot}$ 与 $X_{\cdot j}$ 固定，则 $V_{ij}^{(\alpha)}(1 \leqslant i \leqslant I-1,\ 1 \leqslant j \leqslant J)$ 应满足以下条件：

$$\max\left(V_{i,j-1}^{(\alpha)}, \sum_{l=1}^{j} X_{\cdot l} - \sum_{k=1}^{i-1} V_{kj}^{(\alpha)} - \sum_{k=i+1}^{I} X_{k\cdot}\right)$$

$$\leqslant V_{ij}^{(\alpha)} \leqslant \min\left(V_{i,j-1}^{(\alpha)} + X_{\cdot j} - \sum_{k=1}^{i-1} (V_{kj}^{(\alpha)} - V_{k,j-1}^{(\alpha)}), X_{i\cdot}\right). \tag{24}$$

证。不失一般性，假定 $X_{kj}^{(\alpha)}$ 已确定 $(1 \leqslant k \leqslant i-1, 1 \leqslant j \leqslant J)$。作一个二维列联表 $\{X_{i'j}'^{(\alpha)}\}(1 \leqslant i' \leqslant 2)$，其中 $X_{1j}'^{(\alpha)} = X_{ij}^{(\alpha)}$，$X_{2j}'^{(\alpha)} = \sum_{k=i+1}^{I} X_{kj}^{(\alpha)}$。于是有

$$X_{1\cdot}'^{(\alpha)} = X_{i\cdot}, X_{2\cdot}'^{(\alpha)} = \sum_{k=i+1}^{I} X_{k\cdot}, X_{\cdot j}'^{(\alpha)} = X_{\cdot j} - \sum_{k=1}^{i-1} X_{kj}^{(\alpha)}(1 \leqslant j \leqslant J). \tag{25}$$

据引理 1 有

$$\max\left(V_{1,j-1}'^{(\alpha)}, \sum_{l=1}^{j} X_{\cdot l}'^{(\alpha)} - X_{2\cdot}'^{(\alpha)}\right) \leqslant V_{1j}'^{(\alpha)} \leqslant \min(V_{1,j-1}'^{(\alpha)} + X_{\cdot j}'^{(\alpha)}, X_{1\cdot}'^{(\alpha)}). \tag{26}$$

事实上应有 $V_{1j}'^{(\alpha)} = V_{ij}^{(\alpha)}$，$V_{1,j-1}'^{(\alpha)} = V_{i,j-1}^{(\alpha)}$。再将(25)代入(26)，即有

$$\max\left(V_{i,j-1}^{(\alpha)}, \sum_{l=1}^{j}\left(X_{\cdot l} - \sum_{k=1}^{i-1} X_{kl}^{(\alpha)}\right) - \sum_{k=i+1}^{I} X_{k\cdot}\right)$$

$$\leqslant V_{ij}^{(\alpha)} \leqslant \min\left(V_{i,j-1}^{(\alpha)} + X_{\cdot j} - \sum_{k=1}^{i-1} X_{kj}^{(\alpha)}, X_{i\cdot}\right). \tag{27}$$

于是根据定义 1 中 $X_{kl}^{(\alpha)}$ 与 $V_{kl}^{(\alpha)}$、$V_{k,l-1}^{(\alpha)}$ 的关系即可得结论。

这样，一个二维列联表可用网络的一条通路来表示。

对于三维列联表 $\{X_{ijk}\}(1 \leqslant i \leqslant I, 1 \leqslant j \leqslant J, 1 \leqslant k \leqslant K)$，每一节点记为

$$ND_{ijk}^{(\theta_{ijk})}(1 \leqslant i \leqslant I, 0 \leqslant j \leqslant J, 1 \leqslant k \leqslant K-1),$$

其中 $[\theta_{ijk}] = n_{ijk}n_{i,j-1,k}\cdots n_{i0k}\cdots n_{101}$ 为节点序列号，且可有以下分解：$[\theta_{ijk}] = n_{ijk}[\theta_{i,j-1,k}]$。对于节点 $ND_{ijk}^{(\theta_{ijk})}$ 和 $ND_{i,j-1,k}^{(\theta_{i,j-1,k})^*}$ 若有 $[\theta_{i,j-1,k}]^* = [\theta_{i,j-1,k}]$，则这两节点间可有连线相联，这些连线就构成了通路。对应于每一个节点和每条道路有唯一确定的一个非负整数值，记为 $V_{ijk}^{(\alpha)}$，且令 $V_{i0k}^{(\alpha)} = 0$。

定义 2. 三维列联表 $\{X_{ijk}\}$ 中第 (i,j,k) 格第 α 种可能的频数

$$X_{ijk}^{(\alpha)} = V_{ijk}^{(\alpha)} - V_{i,j-1,k}^{(\alpha)}.$$

引理 2. 三维列联表 $\{X_{ijk}\}(1 \leqslant i \leqslant I, 1 \leqslant j \leqslant J, 1 \leqslant k \leqslant 2)$中，若 $X_{ij\cdot}$、$X_{i\cdot k}$ 及

$X_{i\cdot k}$ 固定，则 $V_{ij1}^{(a)}$ 满足以下条件

$$\max\left(V_{i,j-1,1}^{(a)},\ \sum_{s=1}^{j} X_{\cdot s1}-\sum_{r=1}^{i-1} V_{rj1}^{(a)}-\sum_{r=i+1}^{I} X_{r\cdot 1},\ \sum_{s=1}^{j}X_{is\cdot}-X_{i\cdot 2}\right)\leqslant V_{ij1}^{(a)}$$

$$\leqslant \min\left(V_{i,j-1,1}^{(a)}+X_{\cdot j1}-\sum_{r=1}^{i-1}\left(V_{rj1}^{(a)}-V_{r,j-1,1}^{(a)}\right), X_{i\cdot 1}, V_{i,j-1,1}^{(a)}+X_{ij\cdot}\right). \quad (28)$$

证. 在 $\{X_{ijk}\}$ 中，若 $X_{is1}^{(a)}(1\leqslant s\leqslant j-1)$ 已确定，则根据定理 1 应有

$$\max\left(V_{i,j-1,1}^{(a)},\ \sum_{s=1}^{j} X_{\cdot s1}-\sum_{r=1}^{i-1} V_{rj1}^{(a)}-\sum_{r=i+1}^{I} X_{r\cdot 1}\right)\leqslant V_{ij1}^{(a)}. \quad (29)$$

又因为

$$\sum_{s=1}^{j}\left(X_{is\cdot}-X_{is1}^{(a)}\right)\leqslant X_{i\cdot 2}, \sum_{s=1}^{j} X_{is\cdot}-\sum_{s=1}^{j}\left(V_{is1}^{(a)}-V_{i,s-1,1}^{(a)}\right)\leqslant X_{i\cdot 2},$$

$$\sum_{s=1}^{j} X_{is\cdot}-V_{ij1}^{(a)}\leqslant X_{i\cdot 2},$$

所以

$$V_{ij1}^{(a)}\geqslant \sum_{s=1}^{j} X_{is\cdot}-X_{i\cdot 2}. \quad (30)$$

据(29)和(30)有

$$\max\left(V_{i,j-1,1}^{(a)},\ \sum_{s=1}^{j} X_{\cdot s1}-\sum_{r=1}^{i-1} V_{rj1}^{(a)}-\sum_{r=i+1}^{I} X_{r\cdot 1},\ \sum_{s=1}^{j}X_{is\cdot}-X_{i\cdot 2}\right)\leqslant V_{ij1}^{(a)}. \quad (31)$$

另一方面，据定理 1 应有

$$V_{ij1}^{(a)}\leqslant \min\left(V_{i,j-1,1}^{(a)}+X_{\cdot j1}-\sum_{r=1}^{i-1}\left(V_{rj1}^{(a)}-V_{r,j-1,1}^{(a)}\right), X_{i\cdot 1}\right). \quad (32)$$

又因为 $X_{ij1}^{(a)}\leqslant X_{ij\cdot}$，$V_{ij1}^{(a)}-V_{i,j-1,1}^{(a)}\leqslant X_{ij\cdot}$，所以

$$V_{ij1}^{(a)}\leqslant V_{i,j-1,1}^{(a)}+X_{ij\cdot}. \quad (33)$$

据(32)和(33)有

$$V_{ij1}^{(a)}\leqslant \min\left(V_{i,j-1,1}^{(a)}+X_{\cdot j1}-\sum_{r=1}^{i-1}\left(V_{rj1}^{(a)}-V_{r,j-1,1}^{(a)}\right), X_{i\cdot 1}, V_{i,j-1,1}^{(a)}+X_{ij\cdot}\right). \quad (34)$$

由(31)和(34)即可得结论.

定理 2. 三维列联表$\{X_{ijk}\}(1\leqslant i\leqslant I, 1\leqslant j\leqslant J,\ 1\leqslant k\leqslant K)$中，若 $X_{ij\cdot}$、$X_{i\cdot k}$ 及 $X_{\cdot jk}$ 固定，则 $V_{ijk}^{(a)}$ 满足以下条件

$$\max\left(V_{i,j-1,k}^{(a)}, \sum_{s=1}^{j} X_{\cdot sk}-\sum_{r=1}^{i-1} V_{rjk}^{(a)}-\sum_{r=i+1}^{I} X_{r\cdot k},\right.$$

$$\left.\sum_{s=1}^{j} X_{is\cdot}-\sum_{t=1}^{k-1} V_{ijt}^{(a)}-\sum_{t=k+1}^{K} X_{i\cdot t}\right)\leqslant V_{ijk}^{(a)}$$

$$\leqslant \min\left(V_{i,j-1,k}^{(a)}+X_{\cdot jk}-\sum_{r=1}^{i-1}\left(V_{rjk}^{(a)}-V_{r,j-1,k}^{(a)}\right), X_{i\cdot k},\right.$$

$$V_{i,j-1,k}^{(a)}+X_{ij.}-\sum_{t=1}^{k-1}(V_{ijt}^{(a)}-V_{i,j-1,t}^{(a)})\Big). \tag{35}$$

证．不失一般性，假定 $X_{ijt}^{(a)}$ 已确定 $(1\leqslant i\leqslant I,\ 1\leqslant j\leqslant J,\ 1\leqslant t\leqslant k-1)$．作一个三维列联表 $\{X_{ijk'}^{\prime(a)}\}(1\leqslant k'\leqslant 2)$，其中 $X_{ij1}^{\prime(a)}=X_{ijk}^{(a)}$，$X_{ij2}^{\prime(a)}=\sum\limits_{t=k+1}^{k}X_{ijt}^{(a)}$．于是有

$$X_{.j1}^{\prime(a)}=X_{.jk},\ X_{i.1}^{\prime(a)}=X_{i.k},\ X_{i.2}^{\prime(a)}=\sum_{t=k+1}^{K}X_{i.t},\ X_{.j2}^{\prime(a)}=\sum_{t=k+1}^{K}X_{.jt},$$

$$X_{ij.}^{\prime(a)}=X_{ij.}-\sum_{t=1}^{k-1}X_{ijt}^{(a)}(1\leqslant i\leqslant I,\quad 1\leqslant j\leqslant J). \tag{36}$$

据引理 2 有

$$\max\Big(V_{i,j-1,1}^{\prime(a)},\ \sum_{s=1}^{j}X_{.s1}^{\prime(a)}-\sum_{r=1}^{i-1}V_{rj1}^{\prime(a)}-\sum_{r=i+1}^{I}X_{r.1}^{\prime(a)},\sum_{s=1}^{j}X_{is.}^{\prime(a)}-X_{i.2}^{\prime(a)}\Big)\leqslant V_{ij1}^{\prime(a)}$$

$$\leqslant\min\Big(V_{i,j-1,1}^{\prime(a)}+X_{.j1}^{\prime(a)}-\sum_{r=1}^{i-1}(V_{rj1}^{\prime(a)}-V_{r,j-1,1}^{\prime(a)}),X_{i.1}^{\prime(a)},V_{i,j-1,1}^{\prime(a)}+X_{ij.}^{\prime(a)}\Big). \tag{37}$$

事实上应有 $V_{ij1}^{\prime(a)}=V_{ijk}^{(a)}$，$V_{i,j-1,1}^{\prime(a)}=V_{i,j-1,k}^{(a)}$．再将(36)代入(37)即得结论．

这样，在边际和 $X_{ij.}$、$X_{i.k}$ 及 $X_{.jk}$ 固定的条件下，一个三维列联表可用网络的一条通路来表示．

对于结果 2，其中 $X_{i..}$，$X_{.j.}$，$X_{..k}$ 是固定的，那么根据定理 1 分别可由 $X_{i..}$和 $X_{.j.}$ 得到 $X_{ij.}^{(a)}(1\leqslant a\leqslant A)$，由 $X_{i..}$和 $X_{..k}$ 得到 $X_{i.k}^{(b)}(1\leqslant b\leqslant B)$以及由 $X_{.j.}$和 $X_{..k}$ 得到 $X_{.jk}^{(c)}(1\leqslant c\leqslant C)$，其中 a、b、c 均为正整数．适当组合 $X_{ij.}^{(a)}$，$X_{i.k}^{(b)}$，$X_{.jk}^{(c)}$ 就可以得到满足 $X_{i..}$；$X_{.j.}$，$X_{..k}$ 固定条件的所有可能的列联表．每一种组合对应有一个网络，而每个网络中每条通路即表示一个可能的列联表．

对于结果 3，类似地也可得到若干个网络，每个网络中每条通路表示一个可能的列联表．

§4. 假设检验中概率值的计算

根据 §2 所述结果 1，在二维列联表中，在对二变量独立的原假设作检验时，为计算概率值 P，选择 $\{X'_{ij}\}$ 的条件为

$$\prod_{ij}\frac{X'_{ij}!}{X^*_{ij}!}\geqslant 1. \tag{38}$$

于是当

$$\prod_{i}^{I^*}\prod_{j}^{J^*}X'_{ij}!\geqslant\prod_{ij}X^*_{ij}!(1\leqslant I^*<I,1\leqslant J^*\leqslant J)$$

成立时，显然有 $\prod\limits_{ij}\frac{X'_{ij}!}{X^*_{ij}!}\geqslant 1$．可停止向前搜索，与相应的节点 $ND_{I^*J^*}^{[\theta_{I^*J^*}]}$ 连接的道

路都不必再考虑. 这样由最后余下的各道路相应的表可算得 $P'=\sum f^*(\{X''_{ij}\})$, 其中 $\prod_{ij}(X''_{ij}!/X^*_{ij}!)<1$. 由此可得

$$P=1-P'. \tag{39}$$

同样,对于 §2 所述结果 2 和 3,在三维列联表中,对原假设作检验时, 为计算概率值 P,选择 $\{X'_{ijk}\}$ 的条件为

$$\prod_{ijk}\frac{X'_{ijk}!}{X^*_{ijk}!}\geqslant 1. \tag{40}$$

于是也只要计算 $P'=\sum f^*(\{X''_{ijk}\})$, 其中 $\prod_{ijk}(X''_{ijk}!/X^*_{ijk}!)<1$. 由此即得

$$P=1-P'. \tag{41}$$

§5. 讨　论

为便于讨论,先作以下定义.

定义 3. 在产生二维列联表的网络中,对于节点 $ND_{ij}^{[\theta_{ij}]}$ 有

$$TN_{ij}^{[\theta_{ij}]}=\begin{cases}\sum TN_{i,j+1}^{[\theta_{i,j+1}]^*} & (1\leqslant i\leqslant I-1,\ 0\leqslant j\leqslant J-1),\\ \sum TN_{i+1,1}^{[\theta_{i+1,1}]^*} & (1\leqslant i\leqslant I-2, j=J),\end{cases}$$

其中求和是对所有符合 $[\theta_{i,j+1}]^*=n_{i,j+1}[\theta_{ij}]$ 的 $ND_{i,j+1}^{[\theta_{i,j+1}]^*}$ 进行的.

可见 $TN_{ij}^{[\theta_{ij}]}$ 即为网络中与节点 $ND_{ij}^{[\theta_{ij}]}$ 相连接的各道路上 $ND_{ij}^{[\theta_{ij}]}$ 以后的所有节点数.

根据 §4 中停止向前搜索的原则,若对于某节点 $ND_{i^*j^*}^{[\theta_{i^*j^*}]}$ 有

$$\prod_{i}^{i^*}\prod_{j}^{j^*}X'_{ij}!\geqslant\prod_{ij}X^*_{ij}!,$$

则不必求出由 $ND_{i^*j^*}^{[\theta_{i^*j^*}]}$ 出发的新节点及相应的各种数值. 于是可少做 $TN_{i^*j^*}^{[\theta_{i^*j^*}]}$ 轮运算(求出节点,相应的 $V_{ij}^{(\theta)}$ 值和 $X_{ij}^{(\theta)}$ 值以及 $\Pi X_{ij}!$ 值).

现以一计算实例加以说明. 现观察到一、二维列联表 $\{X^*_{ij}\}$ $\begin{matrix}2&2&0\\1&1&1\\1&0&1\end{matrix}$, 其边际和为 $X_{1.}=4$, $X_{2.}=3$, $X_{3.}=2$, $X_{.1}=4$, $X_{.2}=3$, $X_{.3}=2$. 在行、列边际和固定条件下,欲检验原假设: 二变量相互独立. 根据前述方法可得网络,列于图 1.

图中 $*$ 表示搜索于该节点停止,---表示不再搜索的道路, 连线上数字表示列联表相应格子中的数值.

根据此网络,所有可能的列联表为 39 个. 而因为 $\prod_i\prod_j X^*_{ij}!=4$, 所以根据停止

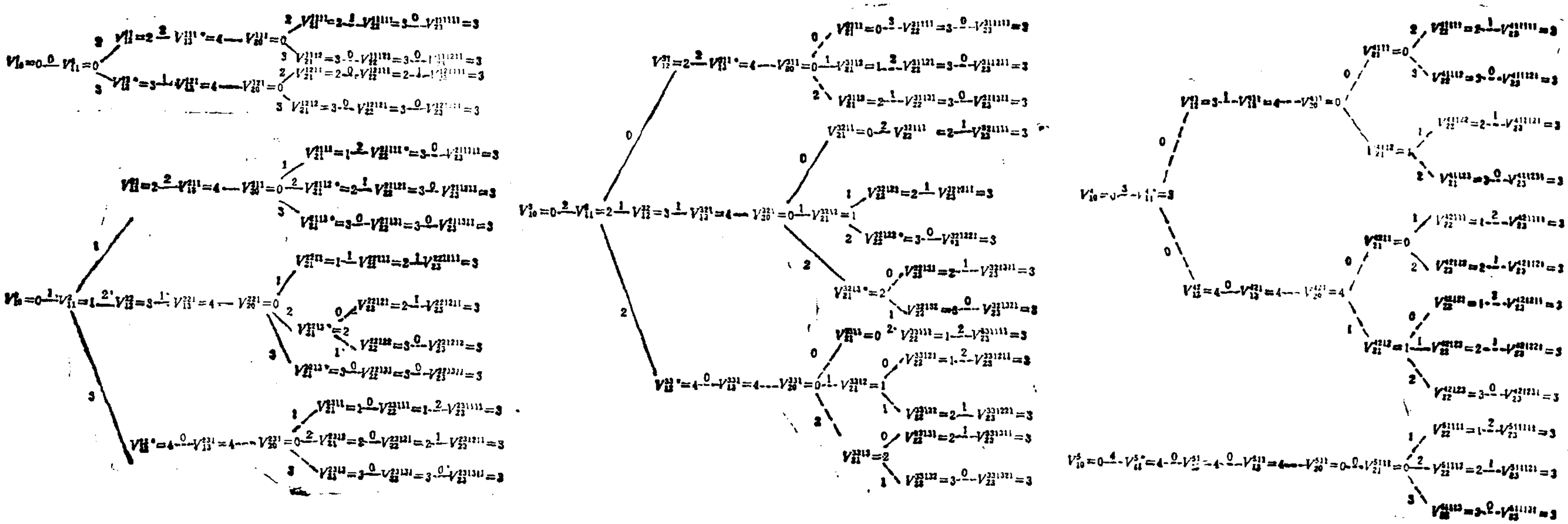

图1　产生边际和为 $X_{1\cdot}=4, X_{2\cdot}=3, X_{3\cdot}=2, X_{\cdot 1}=4, X_{\cdot 2}=3, X_{\cdot 3}=2$ 的所有二维列联表(3×3)的网络

搜索原则，为作确切概率检验，实际上只需求出一张表的确切概率，求其它各表的过程都先后在中途停止。该表的确切概率值为 $\frac{4!\ 3!\ 2!\ 4!\ 3!\ 2!}{2!\ 9!} \doteq 0.114285714$。于是，对原假设作检验，算得的概率值为 $P = 1 - 0.114285714 = 0.885714286$。据此，接受原假设，二变量相互独立。

至于三维列联表的情况可作类似讨论，在此不作赘述。

本文承史秉璋教授审阅指正，谨表谢意。

参考文献

[1] C. R. Rao, Linear Statistical Inferences and Its Applications, John Wiley and Sons, New York, 1973.
[2] Y. M. M. Bishop, S. E. Fienberg, P. M. Holland, Discrete Multivariate Analysis, Theory and Practice. MIT Press, Cambridge, Mass. 1975.
[3] S. E. Fienberg, The Analysis of Cross-Classified Cate gorical Data. 2nd ed. MIT Press, Cambridge Mass, 1980.