対数線形モデルに基づく分割表解析について

対数線形モデルに基づく分割表解析について

鹿児島大学大学院 理工学研究科
数理情報科学専攻 博士前期課程

鬼塚 剛生

2012年2月6日

序文

分割表は，集団データを２つ以上の属性のそれぞれの水準で分類し表でまとめたものであり，属性間の関連を調べることが分割表解析の目的である．分割表解析は医学・心理学・社会学・その他様々な学問において重要な役割を担う．本論文では，属性の影響について解釈しやすい対数線形モデルに焦点を当て，それに基づいた分割表解析の理論と実際のデータ解析の例をまとめている．

第１章では，２元分割表をもとに，分割表解析の土台となる部分について記述している．

第２章では，対象を３元分割表とし，第１章の議論を拡張している．より複雑になったモデル等について記述している．

第３章では，対象を多元分割表とし，一般的な解析について記述している．

分割表の対数線形モデルには数多く候補モデルが存在しうる．そこで第４章では，それらのモデルの選別に有用な対数線形モデルの選択法について記述している．

第５章では，グラフ理論を用いた対数線形モデルのグラフ表現について記述している．

また，第３章を除くすべての章末では，統計解析向けプログラミング言語 R を用いたデータ解析の実行例を記載した．

謝辞

指導教員である青木敏先生には，学部生の頃から大変熱心かつ丁寧なご指導を頂きました．本論文の作成においても，ご多忙の中，ご指導・助言を頂きました．心から感謝申し上げます．

学部・大学院でご指導頂きました数理情報科学科の先生方に深く感謝申し上げます．

大学院にて，同じ研究室でゼミの時間を共にし，励みあった M1 の上籠君，学部４年時に共に学んだ中山君，古瀬君，別府君に感謝致します．

最後に，数理情報科学科の級友たち，学科事務室の方々，応援してくれた家族，学生時代に関わってくださったすべての方々に感謝致します．

目 次

導入

ここでは，分割表の問題を考える際に必要もしくは有効な道具について，最低限の記述をしておく．

分割表

下の表は女性 500 人，男性 600 人，計 1100 人にあるアンケートを行い, その結果を表の形式でまとめたものである．

	賛成	反対	計
女性	309	191	500
男性	319	281	600
計	628	472	1100

上の表のように，ある集団データをいくつかのカテゴリにより分類し，結果をまとめたものを**分割表**（contingency table）という．

離散型確率変数

二項分布

成功の確率が p，失敗の確率が $1-p$ である試行を独立に N 回繰り返したとき，成功回数 X は離散型確率変数であり，確率関数は

$$\Pr(X=n)=\binom{N}{n}p^n(1-p)^{N-n}$$

となる．これをパラメータ N と p の**二項分布**といい，$\mathrm{Bin}(N,p)$ と表す．平均と分散はそれぞれ $\mathrm{E}(X)=Np, \mathrm{Var}(X)=Np(1-p)$ である．

多項分布

結果が $A_1,\ldots,A_I$ の全 I 通りある試行で，各結果が起こる確率を $p_i(i=1,\ldots,I;\sum_{i=1}^{I}p_i=1)$ とする．その試行を独立に N 回繰り返すとき, 結果 $(A_1,\ldots,A_I)$ の起こる回数 $(X_1,\ldots,X_I)$ の同時確率関数は,

$$\Pr(X_1=n_1,...,X_I=n_I)=\frac{N!}{\prod_{i=1}^{I}n_i!}\prod_{i=1}^{I}p_i^{n_i}$$

となる（ただし n_i は非負整数）．これをパラメータ $N, p_1, \ldots, p_I$ の**多項分布**といい，$\mathrm{Mult}(N, p_1, \ldots, p_I)$ と表す．ここで個々の結果の生起回数については，$X_i \sim \mathrm{Bin}(N, p_i)$ が成り立つ．これより $\mathrm{E}(X_i) = Np_i, \mathrm{Var}(X_i) = Np_i(1-p_i)$ である．

積多項分布

I 個の母集団が存在し，それぞれがある多項分布に互いに独立に従う，すなわち

$$\begin{cases} \boldsymbol{X}_i = (X_{i1}, \ldots, X_{is_i}) \sim \mathrm{Mult}(N_i, p_{i1}, \ldots, p_{is_i}),\ i = 1, \ldots, I \\ \boldsymbol{X}_1 \perp\!\!\!\perp \boldsymbol{X}_2 \perp\!\!\!\perp \cdots \perp\!\!\!\perp \boldsymbol{X}_I \end{cases}$$

であるとする $\left(\sum_{j=1}^{s_i} p_{ij} = 1\right)$．このとき，$\boldsymbol{X} = (X_{11}, \ldots, X_{1s_1}, \ldots, X_{I1}, \ldots, X_{Is_I})$ は**積多項分布**（product-multinomial distribution）に従うという．$\boldsymbol{X}_1, \ldots, \boldsymbol{X}_I$ は互いに独立なので，$\boldsymbol{X}$ の同時確率関数は，

$$\begin{aligned} \Pr(\boldsymbol{X} = \boldsymbol{n}) &= \prod_{i=1}^{I} \Pr(X_{i1} = n_{i1}, \ldots, X_{is_i} = n_{is_i}) \\ &= \prod_{i=1}^{I} \left[\frac{N_i!}{\prod_{j=1}^{s_i} n_{ij}!} \prod_{j=1}^{s_i} p_{ij}^{n_{ij}} \right] \end{aligned}$$

となる．

オッズ

定義 0.0.1. ある事象 A の起こる確率を p とするとき，

$$\frac{p}{1-p}$$

のことを事象 A の**オッズ**（odds）とよぶ．

オッズがとりうる値は $0 < \mathrm{odds} < \infty$ である．例えば，あるくじが当たる確率が 0.2 だとすると，そのオッズは $0.2/0.8 = 0.25$ となる．オッズが大きければ確率も大きい，また，オッズが 0 に近ければ確率も小さい．

オッズを用いた統計量に，オッズ比がある．

定義 0.0.2. 2 つの群があるとする．ある事象について，第 1 群で起こる確率を p_1，第 2 群で起こる確率を p_2 とするとき，2 つの群のオッズの比，すなわち

$$\frac{p_1/(1-p_1)}{p_2/(1-p_2)} = \frac{p_1(1-p_2)}{p_2(1-p_1)}$$

を**オッズ比**とよぶ．

オッズ比は，グループ間の比較に用いられ，分割表解析において有効な道具となる．

第1章　2元分割表

分割表は，ある個体の集団をいくつかの属性にしたがい分類し，それぞれの属性の水準の組合せに分類された個体数を表の形で表したものである．例えば，あるクラスの学生を性別（男・女）と血液型（A・B・O・AB）で分類するならば，2×4 分割表が得られる．この場合は，属性が2つであるから**2元分割表**とよぶ．本章では，この2元分割表について記述する．

I 個の水準をもつ属性 A と，J 個の水準をもつ属性 B があるとする．必ず2つの属性のいずれかの水準に分類される個体の集団を考えると，個体数（度数）は $I \times J$ 個のセルを備える $I \times J$ 分割表で表わすことができる．項目 A を行に，B を列に設定し，水準 (A_i, B_j) に分類される度数を n_{ij} とおくと，度数 n_{ij} の $I \times J$ 分割表は下のようにまとめられる．

		属性 B				
		B_1	B_2	$\cdots$	B_J	計
属性 A	A_1	n_{11}	n_{12}	$\cdots$	n_{1J}	$n_{1\cdot}$
	A_2	n_{21}	n_{22}	$\cdots$	n_{2J}	$n_{2\cdot}$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
	A_I	n_{I1}	n_{I2}	$\cdots$	n_{IJ}	$n_{I\cdot}$
	計	$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot J}$	$n_{\cdot\cdot}$

ここで,「$\cdot$」はある添字について和をとった，という意味を表す．この表記により，上の表のように度数に関する第 i 行の行和，第 j 列の列和，総和をそれぞれ

$$n_{i\cdot} = \sum_{j=1}^{J} n_{ij},\ n_{\cdot j} = \sum_{i=1}^{I} n_{ij},\ n_{\cdot\cdot} = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}$$

と定義する．

上は，度数 n_{ij} に関する表であったが，

$$\begin{array}{rcl} X_{ij} & : & \text{セル } (i,j) \text{ の度数を表す確率変数} \\ p_{ij} & : & \text{セル } (i,j) \text{ に属す確率} \\ m_{ij} & : & X_{ij}\text{の期待値（}\textbf{期待度数}\text{）} \end{array}$$

についても上と同様な表が与えられる．

1.1　積多項サンプリングと多項サンプリング

見た目は同じ $I \times J$ 分割表でも，サンプリングの違いで表の性質が変わる．

1.1.1 積多項サンプリング

例 1.1.1. 男性 $n_{1\cdot}$ 人，女性 $n_{2\cdot}$ 人に賛成か反対かで答えられる質問をする．男性が賛成と答える確率を p_1，女性が賛成と答える確率を p_2 とする．そして賛成と答える男性人数を示す確率変数，女性人数を示す確率変数をそれぞれ X_1, X_2 とする．以上のことをまとめると次のようになる．

	賛成	反対	計
男性	p_1	$1-p_1$	1
女性	p_2	$1-p_2$	1

	賛成	反対	計
男性	X_1	$n_{1\cdot}-X_1$	$n_{1\cdot}$
女性	X_2	$n_{2\cdot}-X_2$	$n_{2\cdot}$

このとき，調査結果である X_1, X_2 は，それぞれ独立に $\mathrm{Bin}(n_{1\cdot}, p_1), \mathrm{Bin}(n_{2\cdot}, p_2)$ に従うと考えられる．

この例のように, 行和が固定され，それぞれの行が独立で且つ多項分布に従うとき，その表を行和固定の積多項サンプリング表といい，モデルを**積多項サンプリングモデル**という．

	B_1	$\cdots$	B_J	計
$\boldsymbol{X}_1$	X_{11}	$\cdots$	X_{1J}	$n_{1\cdot}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
$\boldsymbol{X}_I$	X_{I1}	$\cdots$	X_{IJ}	$n_{I\cdot}$

	B_1	$\cdots$	B_J	計
$\boldsymbol{p}_1$	p_{11}	$\cdots$	p_{1J}	1
$\vdots$	$\vdots$		$\vdots$	$\vdots$
$\boldsymbol{p}_I$	p_{I1}	$\cdots$	p_{IJ}	1

$$\boldsymbol{X}_i \sim \mathrm{Mult}(n_{i\cdot}, \boldsymbol{p}_i),\ i = 1, \ldots, I,\ \boldsymbol{X}_1 \perp\!\!\!\perp \ldots \perp\!\!\!\perp \boldsymbol{X}_I$$

列和が固定の場合も同様である．

このような表に対する興味のひとつに,「母集団によって結果に違いがあるかどうか」がある．先ほどの例 1.1.1 でいえば，性別によって回答結果に差があるかどうかである．もし差があるならば $p_1 \neq p_2$ であるし，差がなければ $p_1 = p_2$ となるはずである．本論文では，行和固定の積多項サンプリング表に対しては，母集団に差がないことを帰無仮説 H_0，差があることを対立仮説 H_1 に設定した検定問題，つまり

$$\begin{aligned} \mathrm{H}_0 &: p_{1j} = p_{2j} = \cdots = p_{Ij},\ j = 1, \ldots, J \\ \mathrm{H}_1 &: \mathrm{H}_0\text{は偽（}p\text{ に関する制約は無し）} \end{aligned} \tag{1.1}$$

という検定問題を考える．

1.1.2 多項サンプリング

例 1.1.2. $n_{\cdot\cdot}$ 人に賛成か反対かで答えられる 2 種類の質問をする．この場合，2×2 の計 4 通りの回答がある．それぞれの回答をする確率 p_{ij}（$\sum_{ij} p_{ij} = 1$）とその人数を示す確率変数 X_{ij} をそれぞれ次のようにおく．

		質問 2		
$\boldsymbol{p}$		賛成	反対	計
質問 1	賛成	p_{11}	p_{12}	$p_{1\cdot}$
	反対	p_{21}	p_{22}	$p_{2\cdot}$
		$p_{\cdot 1}$	$p_{\cdot 2}$	1

		質問 2		
$\boldsymbol{X}$		賛成	反対	計
質問 1	賛成	X_{11}	X_{12}	$X_{1\cdot}$
	反対	X_{21}	X_{22}	$X_{2\cdot}$
		$X_{\cdot 1}$	$X_{\cdot 2}$	$n_{\cdot\cdot}$

このとき $\boldsymbol{X} = (X_{11}, X_{12}, X_{21}, X_{22})$ は $\mathrm{Mult}(n_{\cdot\cdot}, p_{11}, p_{12}, p_{21}, p_{22})$ に従うと考えられる．

この例のように，総和が固定されており，セル確率変数ベクトル $\boldsymbol{X}$ が $\mathrm{Mult}(n_{\cdot\cdot}, \boldsymbol{p})$ に従うとき，その表を多項サンプリング表といい，モデルを**多項サンプリングモデル**という．

$\boldsymbol{X}$	B_1	$\cdots$	B_J	計
A_1	X_{11}	$\cdots$	X_{1J}	$X_{1\cdot}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
A_I	X_{I1}	$\cdots$	X_{IJ}	$X_{I\cdot}$
計	$X_{\cdot 1}$	$\cdots$	$X_{\cdot J}$	$n_{\cdot\cdot}$

$\boldsymbol{p}$	B_1	$\cdots$	B_J	計
A_1	p_{11}	$\cdots$	p_{1J}	$p_{1\cdot}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
A_I	p_{I1}	$\cdots$	p_{IJ}	$p_{I\cdot}$
計	$p_{\cdot 1}$	$\cdots$	$p_{\cdot J}$	1

$$(X_{11}, \ldots, X_{IJ}) \sim \mathrm{Mult}(n_{\cdot\cdot}, (p_{11}, \ldots, p_{IJ}))$$

このような表に対する興味のひとつに，行と列の関連の有無，つまり**独立**かどうかがある．ここで，p_{ij} により定まる**周辺確率**（marginal probability）は，

$$p_{i\cdot} = \sum_{j=1}^{J} p_{ij},\ p_{\cdot j} = \sum_{i=1}^{I} p_{ij}$$

である．行と列が独立であるとは，任意のセル (i, j) の確率 p_{ij} が，該当する行と列の周辺確率の積で書ける，つまり $p_{ij} = p_{i\cdot}p_{\cdot j}$ であることと同値である．本論文では，総和が固定された多項サンプリング表については，行と列が独立であることを帰無仮説 H_0，そうでないことを対立仮説 H_1 に設定した検定問題，つまり，

$$\begin{aligned} &\mathrm{H}_0 \ : p_{ij} = p_{i\cdot}p_{\cdot j}, \quad i = 1, \ldots, I,\ j = 1, \ldots, J \\ &\mathrm{H}_1 \ : \mathrm{H}_0\text{は偽（}p\text{ に関する制約は無し）} \end{aligned} \tag{1.2}$$

という検定問題を考える．

1.2 最尤推定値

分割表解析に必要となる，積多項モデルと多項モデルのもとでの**最尤推定値**（maximum likelihood estimate, MLE）の算出を行う．

はじめに，分割表モデルの MLE の算出に必要な補題を記述しておく．

補題 1.2.1. $f(p_1,\ldots,p_r)=\sum_{i=1}^{r} n_i \log p_i$ とする．
任意の $i=1,\ldots,r$ に対して，n_i, p_i が $n_i>0, 0<p_i<1, p_\cdot=1$ を満たすとき，$f(p_1,\ldots,p_r)$ の最大値は点 $(p_1,\ldots,p_r)=(\hat{p}_1,\ldots,\hat{p}_r), (\hat{p}_i=n_i/n_\cdot)$ にて与えられる．

（証明）

$$\begin{aligned} f(p_1,\ldots,p_r) &= \sum_{i=1}^{r} n_i \log p_i \\ &= \sum_{i=1}^{r-1} n_i \log p_i + n_r \log p_r \\ &= \sum_{i=1}^{r-1} n_i \log p_i + n_r \log \left(1-\sum_{i=1}^{r-1} p_i\right) \end{aligned}$$

偏微分して 0 とおくと，

$$\frac{\partial f}{\partial p_i} = \frac{n_i}{p_i} - \frac{n_r}{1-\sum_{i=1}^{r-1} p_i} = 0 \quad (i=1,\ldots,r-1)$$

である．これより，

$$\frac{n_i}{p_i} = \frac{n_r}{1-\sum_{i=1}^{r-1} p_i} = \frac{n_r}{p_r} \quad (i=1,\ldots,r-1)$$

である．ここで $n_i/p_i=K$ とすると，$n_i=Kp_i$ $(i=1,\ldots,r)$ であるから，

$$\begin{aligned} n_\cdot = \sum_{i=1}^{r} n_i = \sum_{i=1}^{r} Kp_i = K \\ \therefore \quad n_\cdot = K \end{aligned}$$

となる．これより

$$\begin{aligned} \frac{n_i}{p_i} = K = n_\cdot \\ \therefore \hat{p}_i = \frac{n_i}{n_\cdot} \end{aligned}$$

となる． □

次に，各モデルごとに MLE を算出する．ただし，$n_{ij}>0$ を仮定しておく．

積多項サンプリング

ここでは，同じ J 個のカテゴリに分かれる I 個の母集団の積多項サンプリングモデル，つまり

$$\begin{cases} \boldsymbol{X}_i = (X_{i1},\ldots,X_{iJ}) \sim \mathrm{Mult}(N_i, p_{i1},\ldots,p_{iJ}),\ i=1,\ldots,I \\ \boldsymbol{X}_1 \perp\!\!\!\perp \cdots \perp\!\!\!\perp \boldsymbol{X}_I \end{cases}$$

について考える．導入の定義より，同時確率関数は

$$\Pr(\boldsymbol{X}=\boldsymbol{n})=\prod_{i=1}^{I}\left[\frac{n_{i\cdot}!}{\prod_{j=1}^{J}n_{ij}!}\prod_{j=1}^{J}p_{ij}^{n_{ij}}\right]$$

である．これを $\boldsymbol{p}$ の関数とみなす**尤度関数** $L(\boldsymbol{p})$ は，

$$L(\boldsymbol{p})=\prod_{i=1}^{I}\left[\frac{n_{i\cdot}!}{\prod_{j=1}^{J}n_{ij}!}\prod_{j=1}^{J}p_{ij}^{n_{ij}}\right]$$

である．

まず，H_1 のもと，つまり $\boldsymbol{p}$ に制約がないときの MLE を求める．

$\boldsymbol{p}$ の MLE は尤度関数 $L(\boldsymbol{p})$ を最大にする点 $\hat{\boldsymbol{p}}$ である．$L(\boldsymbol{p})$ の対数をとると，

$$\begin{aligned}
\log L(\boldsymbol{p}) &= \log\prod_{i=1}^{I}\left[\frac{n_{i\cdot}!}{\prod_{j=1}^{J}n_{ij}!}\prod_{j=1}^{J}p_{ij}^{n_{ij}}\right] \\
&= \sum_{i=1}^{I}\left[\log\left(\frac{n_{i\cdot}!}{\prod_{j=1}^{J}n_{ij}!}\prod_{j=1}^{J}p_{ij}^{n_{ij}}\right)\right] \\
&= \sum_{i=1}^{I}\left[\log n_{i\cdot}! - \log\prod_{j=1}^{J}n_{ij}! + \log\prod_{j=1}^{J}p_{ij}^{n_{ij}}\right] \\
&= \sum_{i=1}^{I}\left[\log n_{i\cdot}! - \sum_{j=1}^{J}\log n_{ij}! + \sum_{j=1}^{J}n_{ij}\log p_{ij}\right]
\end{aligned}$$

となる．これを最大にするには $\boldsymbol{p}$ に依存していない部分は無視できるので，

$$\ell(\boldsymbol{p})=\sum_{i=1}^{I}\sum_{j=1}^{J}n_{ij}\log p_{ij}$$

の最大化のみで十分である．$\ell(\boldsymbol{p})$ の最大値は $\sum_{j=1}^{J}n_{ij}\log p_{ij}$ の全ての項を最大にすることで得られる．ここで先の補題 1.2.1 より，これを最大にする点 $\hat{\boldsymbol{p}}$ は

$$\hat{\boldsymbol{p}}\ ;\ \hat{p}_{ij}=\frac{n_{ij}}{n_{i\cdot}}$$

である．以上のように確率の MLE が得られた．H_1 のもとでの期待度数 m_{ij} の MLE は，最尤推定量の不変性により，

$$\hat{m}_{ij}=n_{i\cdot}\hat{p}_{ij}=n_{i\cdot}\frac{n_{ij}}{n_{i\cdot}}=n_{ij}$$

のように得られる．

帰無仮説 $\mathrm{H}_0: p_{1j}=\ldots=p_{Ij},\ j=1,\ldots,J$ のもとでは，先の MLE とは異なるものが得られる．任意の j に対して，$\pi_j=p_{1j}=\cdots=p_{Ij}$ とする．そのときの対数尤度関数は，

$$\log L(\boldsymbol{p})=\sum_{i=1}^{I}\left[\log n_{i\cdot}! - \sum_{j=1}^{J}\log n_{ij}! + \sum_{j=1}^{J}n_{ij}\log\pi_j\right]$$

となる．前と同様に $\boldsymbol{p}$ に依存しない部分は無視し，

$$\ell(\boldsymbol{p}) = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\log\pi_j = \sum_{j=1}^{J} n_{\cdot j}\log\pi_j$$

の最大化のみ考えればよい．補題 1.2.1 より確率の MLE は，

$$\hat{p}_{ij}^{(0)} = \hat{\pi}_j = \frac{n_{\cdot j}}{n_{\cdot\cdot}}$$

となる．ここで上添字 (0) は，仮説 H_0 のもとでの MLE であることを示す．H_0 のもとでの期待度数の MLE は，

$$\hat{m}_{ij}^{(0)} = n_{i\cdot}\hat{p}_{ij}^{(0)} = n_{i\cdot}\frac{n_{\cdot j}}{n_{\cdot\cdot}}$$

である．

多項サンプリング

多項サンプリングモデルの場合を考える．第 (i,j) セルの確率を p_{ij} とすると，

$$\boldsymbol{X} = (X_{11},\ldots,X_{IJ}) \sim \mathrm{Mult}(n_{\cdot\cdot}, p_{11},\ldots,p_{IJ})$$

である．定義より同時確率関数は

$$\Pr(\boldsymbol{X}=\boldsymbol{n}) = \frac{n_{\cdot\cdot}!}{\prod_{i=1}^{I}\prod_{j=1}^{J} n_{ij}!}\prod_{i=1}^{I}\prod_{j=1}^{J} p_{ij}^{n_{ij}}$$

であり，尤度関数は

$$L(\boldsymbol{p}) = \frac{n_{\cdot\cdot}!}{\prod_{i=1}^{I}\prod_{j=1}^{J} n_{ij}!}\prod_{i=1}^{I}\prod_{j=1}^{J} p_{ij}^{n_{ij}}$$

である．

H_1 のもとでの $\boldsymbol{p}$ の MLE を求める．対数尤度関数は，

$$\begin{aligned}
\log L(\boldsymbol{p}) &= \log\left(\frac{n_{\cdot\cdot}!}{\prod_{i=1}^{I}\prod_{j=1}^{J} n_{ij}!}\prod_{i=1}^{I}\prod_{j=1}^{J} p_{ij}^{n_{ij}}\right)\\
&= \log\left(\frac{n_{\cdot\cdot}!}{\prod_{i=1}^{I}\prod_{j=1}^{J} n_{ij}!}\right) + \log\left(\prod_{i=1}^{I}\prod_{j=1}^{J} p_{ij}^{n_{ij}}\right)\\
&= \log\left(\frac{n_{\cdot\cdot}!}{\prod_{i=1}^{I}\prod_{j=1}^{J} n_{ij}!}\right) + \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\log p_{ij}
\end{aligned}$$

となる．積多項のときと同様，$\ell(\boldsymbol{p}) = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\log p_{ij}$ の最大化を考えれば十分である．補題 1.2.1 より尤度関数を最大にする $\boldsymbol{p}$ の MLE は，

$$\hat{\boldsymbol{p}}\ ;\ \hat{p}_{ij} = \frac{n_{ij}}{n_{\cdot\cdot}}$$

となる．これより，H_1 のもとでの期待度数 $\boldsymbol{m}$ の MLE は，$m_{ij} = n_{..}p_{ij}$ より

$$\hat{m}_{ij} = n_{..}\hat{p}_{ij} = n_{..}\frac{n_{ij}}{n_{..}} = n_{ij}$$

である．

帰無仮説 H_0 ： $p_{ij} = p_{i\cdot}p_{\cdot j},\ i = 1,\ldots,I,\ j = 1,\ldots,J$ のもとでは，

$$\log L(\boldsymbol{p}) = \log\left(\frac{n_{..}!}{\prod_{i=1}^{I}\prod_{j=1}^{J} n_{ij}!}\right) + \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\log p_{i\cdot}p_{\cdot j}$$

である．先ほどと同様，$\ell(\boldsymbol{p}) = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\log p_{i\cdot}p_{\cdot j}$ を最大にする $\boldsymbol{p}$ を求める．変形すると，

$$\begin{aligned}
\ell(\boldsymbol{p}) &= \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\mathrm{log}p_{i\cdot}p_{\cdot j} \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}(\mathrm{log}p_{i\cdot} + \mathrm{log}p_{\cdot j}) \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\mathrm{log}p_{i\cdot} + \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\mathrm{log}p_{\cdot j} \\
&= \sum_{i=1}^{I} n_{i\cdot}\mathrm{log}p_{i\cdot} + \sum_{j=1}^{J} n_{\cdot j}\mathrm{log}p_{\cdot j}
\end{aligned}$$

となる．$n_{i\cdot}, n_{\cdot j} > 0, 0 < p_{i\cdot}, p_{\cdot j} < 1, p_{\cdot\cdot} = 1$ であるから，補題 1.2.1 より確率の MLE は

$$\hat{p}_{i\cdot}^{(0)} = \frac{n_{i\cdot}}{n_{..}}, \hat{p}_{\cdot j}^{(0)} = \frac{n_{\cdot j}}{n_{..}}$$
$$\therefore\ \hat{p}_{ij}^{(0)} = \hat{p}_{i\cdot}\hat{p}_{\cdot j} = \left(\frac{n_{i\cdot}}{n_{..}}\right)\left(\frac{n_{\cdot j}}{n_{..}}\right)$$

となる．これより H_0 のもとでの期待度数の MLE は，

$$\hat{m}_{ij}^{(0)} = n_{..}\hat{p}_{ij}^{(0)} = n_{..}\left(\frac{n_{i\cdot}}{n_{..}}\right)\left(\frac{n_{\cdot j}}{n_{..}}\right) = \frac{n_{i\cdot}n_{\cdot j}}{n_{..}}$$

となる．

多項サンプリング表に関していえば，別の考え方で周辺確率の MLE の算出が可能である．多項サンプリング表の確率変数と確率の表は，

$\boldsymbol{X}$	B_1	$\cdots$	B_J	計
A_1	X_{11}	$\cdots$	X_{1J}	$X_{1\cdot}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
A_I	X_{I1}	$\cdots$	X_{IJ}	$X_{I\cdot}$
計	$X_{\cdot 1}$	$\cdots$	$X_{\cdot J}$	$n_{..}$

$\boldsymbol{p}$	B_1	$\cdots$	B_J	計
A_1	p_{11}	$\cdots$	p_{1J}	$p_{1\cdot}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
A_I	p_{I1}	$\cdots$	p_{IJ}	$p_{I\cdot}$
計	$p_{\cdot 1}$	$\cdots$	$p_{\cdot J}$	1

である．ここで行和 $X_{i\cdot}(i=1,\ldots,I)$ に注目すると，

$$(X_{1\cdot},\ldots,X_{I\cdot}) \sim \mathrm{Mult}(n_{\cdot\cdot}, p_{1\cdot},\ldots,p_{I\cdot})$$

である．これより $X_{i\cdot}(i=1,\ldots,I)$ の同時確率関数，尤度はそれぞれ

$$\Pr(X_{i\cdot}=n_{i\cdot}, i=1,\ldots,I) = \frac{n_{\cdot\cdot}!}{\prod_{i=1}^{I} n_{i\cdot}!}\prod_{i=1}^{I} p_{i\cdot}^{n_{i\cdot}}$$

$$L(p_{i\cdot}, i=1,\cdots,I) = \frac{n_{\cdot\cdot}!}{\prod_{i=1}^{I} n_{i\cdot}!}\prod_{i=1}^{I} p_{i\cdot}^{n_{i\cdot}}$$

である．対数尤度は

$$\log L(p_{i\cdot}, i=1,\cdots,I) = \log\left(\frac{n_{\cdot\cdot}!}{\prod_{i=1}^{I} n_{i\cdot}!}\right) + \sum_{i=1}^{I} n_{i\cdot}\log p_{i\cdot}$$

である．$p_{i\cdot}$ に依存する項 $\ell(p_{i\cdot}, i=1,\cdots,I) = \sum_{i=1}^{I} n_{i\cdot}\log p_{i\cdot}$ に対して補題 1.2.1 を適用することで，行の周辺確率 $p_{i\cdot}$ の MLE は

$$\hat{p}_{i\cdot} = \frac{n_{i\cdot}}{n_{\cdot\cdot}}, \quad i=1,\ldots,I$$

となる．同様に，列の周辺確率 $p_{\cdot j}$ の MLE は

$$\hat{p}_{\cdot j} = \frac{n_{\cdot j}}{n_{\cdot\cdot}}, \quad j=1,\ldots,J$$

である．これより，$\mathrm{H}_0 : p_{ij} = p_{i\cdot}p_{\cdot j}$ の下での MLE が $\hat{p}_{ij}^{(0)} = (n_{i\cdot}/n_{\cdot\cdot})(n_{\cdot j}/n_{\cdot\cdot})$ であることがわかる．

本節で行った MLE の算出の方法は，３元以上の分割表のモデルでも同様である．

1.3 モデル検定

1.3.1 検定統計量

ピアソンカイ二乗検定統計量

H_0 に対する代表的な検定統計量のひとつに，ピアソンカイ二乗検定統計量がある．$I\times J$ 分割表データに関して，n_{ij} : 観測度数, $\hat{m}_{ij}$: H_0のもとでの期待度数 とするとき,

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(n_{ij}-\hat{m}_{ij})^2}{\hat{m}_{ij}}$$

を H_0 に対する**ピアソンカイ二乗統計量**（Pearson chi-square test statistic）とよぶ．これは現実の値 n_{ij} と，H_0 のもとで当てはめ値である $\hat{m}_{ij}$ のずれを表わす値であり，分母の $\hat{m}_{ij}$ は尺度係数である．

第 1.1 節の２元分割表の２つのサンプリングモデルにおける帰無仮説 H_0(1.1),(1.2) のもとでは，X^2 は漸近的に $\chi^2_{(I-1)(J-1)}$ 分布（自由度 $(I-1)(J-1)$ のカイ二乗分布）に従うことが知られている．これを利用した検定をカイ二乗適合度検定とよぶ.

尤度比検定統計量

尤度比検定統計量は,

$$G^2 = 2\log\left[\frac{L(\hat{\boldsymbol{p}})}{L(\hat{\boldsymbol{p}}^{(\mathbf{0})})}\right] \tag{1.3}$$

と定義される．ここで，$\hat{\boldsymbol{p}}^{(0)}$ は H_0 のもとでの $\boldsymbol{p}$ の MLE であり，$\hat{\boldsymbol{p}}$ は制約が課されていない場合の $\boldsymbol{p}$ の MLE である．つまり,

$$\begin{aligned} L(\hat{\boldsymbol{p}}^{(\mathbf{0})}):&\ \ \mathrm{H}_0\text{のもとでの尤度（制約があるもとでの尤度）} \\ L(\hat{\boldsymbol{p}}):&\ \ \text{制約がないときの尤度} \end{aligned}$$

である．そもそも尤度関数は，母数 $\boldsymbol{p}$ の尤もらしさを表わす関数であり，$\boldsymbol{p}$ に制約 H_0 を課すと，尤度は下がる．このことから，H_0 を課したときの尤度が十分に小さいとき，その H_0 は正当ではないと判断することにし，尤度の相対的な縮減を $L(\boldsymbol{p})$ の最大値との比，つまり

$$\frac{L(\hat{\boldsymbol{p}}^{(\mathbf{0})})}{L(\hat{\boldsymbol{p}})}$$

として定義し，これを検定統計量にするものが尤度比検定である．便宜上，この尤度比そのものを基準にするのではなく対数尤度比，つまり

$$\log\left[\frac{L(\hat{\boldsymbol{p}}^{(\mathbf{0})})}{L(\hat{\boldsymbol{p}})}\right] = \log L(\hat{\boldsymbol{p}}^{(\mathbf{0})}) - \log L(\hat{\boldsymbol{p}})$$

を利用する．これに -2 をかけたもの，つまり

$$G^2 = 2\log\left[\frac{L(\hat{\boldsymbol{p}})}{L(\hat{\boldsymbol{p}}^{(\mathbf{0})})}\right]$$

を**尤度比検定統計量**（likelihood ratio test statistic）とし，この値が非常に大きくなった場合に H_0 を棄却する．

積多項サンプリング　積多項サンプリングモデルの仮説 H_0(1.1) の G^2 をみる．$\log L(\hat{\boldsymbol{p}}^{(0)}) - \log L(\hat{\boldsymbol{p}})$ の 2 つの項は，それぞれ

$$\begin{aligned} \log L(\hat{\boldsymbol{p}}^{(0)}) &= \sum_{i=I}^{I}\left[\log n_{i\cdot}! - \sum_{j=1}^{J}\log n_{ij}! + \sum_{j=1}^{J} n_{ij}\log\left(\frac{n_{\cdot j}}{n_{\cdot\cdot}}\right)\right] \\ \log L(\hat{\boldsymbol{p}}) &= \sum_{i=I}^{I}\left[\log n_{i\cdot}! - \sum_{j=1}^{J}\log n_{ij}! + \sum_{j=1}^{J} n_{ij}\log\left(\frac{n_{ij}}{n_{i\cdot}}\right)\right] \end{aligned}$$

である（第 1.2 節の結果を利用した）．対数尤度比は,

$$\log L(\hat{\boldsymbol{p}}^{(0)}) - \log L(\hat{\boldsymbol{p}}) = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\log\left(\frac{n_{\cdot j}}{n_{\cdot\cdot}}\right) - \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\log\left(\frac{n_{ij}}{n_{i\cdot}}\right)$$

$$
\begin{aligned}
&= \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\left\{\log\left(\frac{n_{\cdot j}}{n_{\cdot\cdot}}\right) - \log\left(\frac{n_{ij}}{n_{i\cdot}}\right)\right\} \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\log\left(\frac{n_{\cdot j}n_{i\cdot}}{n_{\cdot\cdot}n_{ij}}\right) \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J} \hat{m}_{ij}\log\left(\frac{\hat{m}_{ij}^{(0)}}{\hat{m}_{ij}}\right) \\
&= -\sum_{i=1}^{I}\sum_{j=1}^{J} \hat{m}_{ij}\log\left(\frac{\hat{m}_{ij}}{\hat{m}_{ij}^{(0)}}\right)
\end{aligned}
\tag{1.4}
$$

となる．これに -2 をかけると，

$$
G^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J} \hat{m}_{ij}\log\left(\frac{\hat{m}_{ij}}{\hat{m}_{ij}^{(0)}}\right)
$$

となる．大標本のときの漸近分布

$$
G^2 \sim \chi^2_{(I-1)(J-1)}
$$

を利用して検定を行う．

多項サンプリング　多項サンプリングモデルの仮説 $\mathrm{H}_0(1.2)$ の G^2 をみる．対数尤度はそれぞれ，

$$
\log L(\hat{\boldsymbol{p}}^{(0)}) = \log\left(\frac{n_{\cdot\cdot}!}{\prod_{i=1}^{I}\prod_{j=1}^{J} n_{ij}!}\right) + \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\log\left(\frac{n_{i\cdot}}{n_{\cdot\cdot}}\cdot\frac{n_{\cdot j}}{n_{\cdot\cdot}}\right)
$$

$$
\log L(\hat{\boldsymbol{p}}) = \log\left(\frac{n_{\cdot\cdot}!}{\prod_{i=1}^{I}\prod_{j=1}^{J} n_{ij}!}\right) + \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\log\frac{n_{ij}}{n_{\cdot\cdot}}
$$

である（こちらも第 1.2 節の結果を利用した）．これより尤度比は，

$$
\begin{aligned}
\log L(\hat{\boldsymbol{p}}^{(0)}) - \log L(\hat{\boldsymbol{p}}) &= \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\log\left(\frac{n_{i\cdot}}{n_{\cdot\cdot}}\cdot\frac{n_{\cdot j}}{n_{\cdot\cdot}}\cdot\frac{n_{\cdot\cdot}}{n_{ij}}\right) \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\log\left(\frac{n_{i\cdot}n_{\cdot j}}{n_{\cdot\cdot}n_{ij}}\right) \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J} \hat{m}_{ij}\log\left(\frac{\hat{m}_{ij}^{(0)}}{\hat{m}_{ij}}\right) \\
&= -\sum_{i=1}^{I}\sum_{j=1}^{J} \hat{m}_{ij}\log\left(\frac{\hat{m}_{ij}}{\hat{m}_{ij}^{(0)}}\right)
\end{aligned}
\tag{1.5}
$$

となり，これに -2 をかけた

$$
G^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J} \hat{m}_{ij}\log\left(\frac{\hat{m}_{ij}}{\hat{m}_{ij}^{(0)}}\right)
$$

を検定統計量とする．積多項の場合と同様，大標本のときの漸近分布 $G^2 \sim \chi^2_{(I-1)(J-1)}$ を利用できる．

以上のように，積多項モデルと多項モデルの H_0 に対する G^2 は同じ式になる．また，計算過程である (1.4),(1.5) が同じ式であるため，値も等しくなる．

もともとの尤度比統計量は式 (1.3) のように定義されていたが，分割表解析においては一般的に，いま算出したような，

$$G^2 = 2\sum \hat{\boldsymbol{m}} \log\left(\frac{\hat{\boldsymbol{m}}}{\hat{\boldsymbol{m}}^{(0)}}\right)$$

$\hat{\boldsymbol{m}}$: 制約なしの m の MLE
$\hat{\boldsymbol{m}}^{(0)}$: ある仮説 H_0のもとでの m の MLE

を利用する．何も制約がない場合は，$\hat{\boldsymbol{m}}$=$\boldsymbol{n}$ となるため，ある制約 H_0 に対する検定統計量は

$$G^2 = 2\sum \boldsymbol{n} \log\left(\frac{\boldsymbol{n}}{\hat{\boldsymbol{m}}^{(0)}}\right)$$

となる．

1.3.2 オッズ比

2 つのサンプリングモデルの帰無仮説 H_0(1.1),(1.2) は，それぞれオッズ比を用いた表現に書き換えられる．これにより，これらの検定はオッズ比に関する検定に変換できる．ここでいうオッズ比とは，

$$\frac{p_{ij}/p_{ij'}}{p_{i'j}/p_{i'j'}} = \frac{p_{ij}p_{i'j'}}{p_{ij'}p_{i'j}} \quad (\text{ただし}, i \neq i',\ j \neq j')$$

のことである．

命題 1.3.1. 行和固定の積多項サンプリングモデルにおいて，

$$p_{1j} = \cdots = p_{Ij},\ j = 1, \ldots, J$$
$$\Longleftrightarrow \frac{p_{ij}p_{i'j'}}{p_{ij'}p_{i'j}} = 1, \quad i, i' = 1, \ldots, I,\ j, j' = 1, \ldots, J$$

である．

(証明)
($\Longrightarrow$)

$$\begin{aligned}\frac{p_{ij}p_{i'j'}}{p_{ij'}p_{i'j}} &= \frac{p_{ij}p_{ij'}}{p_{ij'}p_{ij}} \quad (\because \text{仮定より } p_{i'j'} = p_{ij'},\ p_{i'j} = p_{ij}) \\ &= 1\end{aligned}$$

($\Longleftarrow$)
任意の i に対して，$p_{i\cdot} = 1$ である．これより，$p_{\cdot\cdot} = \sum_{i=1}^{I} p_{i\cdot} = I$ である．また，仮定より $p_{ij}p_{i'j'} = p_{ij'}p_{i'j} \cdots (*)$ である．すると，

$$\begin{aligned}
p_{ij} &= p_{ij} \cdot 1 = p_{ij}\frac{p_{\cdot\cdot}}{I} \\
&= \frac{1}{I}p_{ij}\sum_{i'=1}^{I}\sum_{j'=1}^{J} p_{i'j'} = \frac{1}{I}\sum_{i'=1}^{I}\sum_{j'=1}^{J} p_{ij}p_{i'j'} \\
&= \frac{1}{I}\sum_{i'=1}^{I}\sum_{j'=1}^{J} p_{ij'}p_{i'j} \quad (\because (*)) \\
&= \frac{1}{I}\left(\sum_{i'=1}^{I} p_{i'j}\right)\left(\sum_{j'=1}^{J} p_{ij'}\right) \\
&= \frac{1}{I}p_{\cdot j}p_{i\cdot} \\
\therefore p_{ij} &= \frac{1}{I}p_{\cdot j}
\end{aligned}$$

である．これが任意の i, j に対して成り立つ．よって，任意の j に対して $p_{\cdot j}/I = p_{1j} = p_{2j} = \cdots = p_{Ij}$ が成り立つ． □

この命題より，行和固定の積多項サンプリングモデルの $\mathrm{H}_0(1.1)$ の検定は，

$$\mathrm{H}_0 : \frac{p_{ij}p_{i'j'}}{p_{ij'}p_{i'j}} = 1 \text{ for all } i, i'j, j',$$

もしくは

$$\mathrm{H}_0 : \log\left(\frac{p_{ij}p_{i'j'}}{p_{ij'}p_{i'j}}\right) = 0 \text{ for all } i, i'j, j'$$

の検定と同値であることがわかる．

命題 1.3.2. 多項サンプリングモデルにおいて，

$$\begin{aligned}
& p_{ij} = p_{i\cdot}p_{\cdot j},\ i = 1, \ldots, I,\ j = 1, \ldots, J \\
\Longleftrightarrow\ & \frac{p_{ij}p_{i'j'}}{p_{ij'}p_{i'j}} = 1,\ i, i' = 1, \ldots, I,\ j, j' = 1, \ldots, J
\end{aligned}$$

（証明）
($\Longrightarrow$)

$$\frac{p_{ij}p_{i'j'}}{p_{ij'}p_{i'j}} = \frac{(p_{i\cdot}p_{\cdot j})(p_{i'\cdot}p_{\cdot j'})}{(p_{i\cdot}p_{\cdot j'})(p_{i'\cdot}p_{\cdot j})} = 1$$

($\Longleftarrow$)
仮定より $p_{ij}p_{i'j'} = p_{ij'}p_{i'j} \cdots (*)$ である．すると，

$$p_{ij} = p_{ij} \cdot 1 = p_{ij}p_{\cdot\cdot}$$

$$
\begin{aligned}
&= p_{ij} \sum_{i'=1}^{I} \sum_{j'=1}^{J} p_{i'j'} = \sum_{i'=1}^{I} \sum_{j'=1}^{J} p_{ij} p_{i'j'} \\
&= \sum_{i'=1}^{I} \sum_{j'=1}^{J} p_{ij'} p_{i'j} \quad (\because (*)) \\
&= \left(\sum_{i'=1}^{I} p_{i'j} \right) \left(\sum_{j'=1}^{J} p_{ij'} \right) \\
p_{ij} &= p_{\cdot j} p_{i\cdot}
\end{aligned}
$$

と書ける．これより，任意の i, j に対して $p_{ij} = p_{i\cdot} p_{\cdot j}$ が成り立つ． □

この命題より，総和固定の多項サンプリングモデルの $\mathrm{H}_0(1.2)$ の検定も積多項表と同様,

$$
\mathrm{H}_0 : \frac{p_{ij} p_{i'j'}}{p_{ij'} p_{i'j}} = 1 \text{ for all } i, i'j, j'
$$

もしくは

$$
\mathrm{H}_0 : \log \left(\frac{p_{ij} p_{i'j'}}{p_{ij'} p_{i'j}} \right) = 0 \text{ for all } i, i'j, j'
$$

の検定と同値であることがわかる．

積多項表と多項表のどちらの仮説も，同値なオッズ比の検定は同じものとなった．
以下は，オッズ比に関する有益な性質である．

命題 1.3.3.

$$
\begin{aligned}
&\frac{p_{ij} p_{i'j'}}{p_{ij'} p_{i'j}} = 1 \text{ for all } \ i, i', j, j' \\
\iff &\frac{p_{11} p_{ij}}{p_{1j} p_{i1}} = 1 \text{ for all } \ i, j \neq 1
\end{aligned}
$$

(証明)
($\Longrightarrow$)
これは明らかに成立する．
($\Longleftarrow$)
$i, j \neq 1, i' \neq 1$ に対して,

$$
\frac{p_{11} p_{ij}}{p_{1j} p_{i1}} = \frac{p_{11} p_{i'j}}{p_{1j} p_{i'1}} = 1
$$

である．これより,

$$
\begin{aligned}
1 &= \left(\frac{p_{11} p_{ij}}{p_{1j} p_{i1}} \right) \Big/ \left(\frac{p_{11} p_{i'j}}{p_{1j} p_{i'1}} \right) \\
&= \frac{p_{ij} p_{i'1}}{p_{i1} p_{i'j}}
\end{aligned}
$$

となる．

また，$i, j \neq 1, j' \neq 1$ に対して，

$$\frac{p_{11}p_{ij}}{p_{1j}p_{i1}} = \frac{p_{11}p_{ij'}}{p_{1j'}p_{i1}} = 1$$

であるから，

$$\begin{aligned} 1 &= \left(\frac{p_{11}p_{ij}}{p_{1j}p_{i1}}\right) \Big/ \left(\frac{p_{11}p_{ij'}}{p_{1j'}p_{i1}}\right) \\ &= \frac{p_{ij}p_{1j'}}{p_{ij'}p_{1j}} \end{aligned}$$

となる．

さらに，$i, i' \neq 1,\ j, j' \neq 1$ に対して，

$$\begin{aligned} 1 &= \left\{\left(\frac{p_{11}p_{ij}}{p_{1j}p_{i1}}\right)\left(\frac{p_{11}p_{i'j'}}{p_{1j'}p_{i'1}}\right)\right\} \Big/ \left\{\left(\frac{p_{11}p_{ij'}}{p_{1j'}p_{i1}}\right)\left(\frac{p_{11}p_{i'j}}{p_{1j}p_{i'1}}\right)\right\} \\ &= \frac{p_{ij}p_{i'j'}}{p_{ij'}p_{i'j}} \end{aligned}$$

となる．

以上より，任意の i, j に対して $(p_{ij}p_{i'j'})/(p_{ij'}p_{i'j}) = 1$ となる． □

ここまで記述してきたオッズ比を構成する p_{ij} は未知である．検定では推定オッズ比，つまり

$$\frac{\hat{p}_{ij}\hat{p}_{i'j'}}{\hat{p}_{ij'}\hat{p}_{i'j}} = \frac{n_{ij}n_{i'j'}}{n_{ij'}n_{i'j}} \quad \left(\text{行和固定の積多項では}\hat{p}_{ij} = \frac{n_{ij}}{n_{i\cdot}}, \text{多項では}\hat{p}_{ij} = \frac{n_{ij}}{n_{\cdot\cdot}}\right)$$

を利用する．

また，２つのモデルに関しては，期待度数は

$$\text{積多項} : m_{ij} = n_{i\cdot}p_{ij},\ \text{多項} : m_{ij} = n_{\cdot\cdot}p_{ij}$$

であるから，

$$\frac{p_{ij}p_{i'j'}}{p_{ij'}p_{i'j}} = \frac{m_{ij}m_{i'j'}}{m_{ij'}m_{i'j}}$$

とも書け，推定された $\boldsymbol{m}$ を代入した，

$$\frac{\hat{m}_{ij}\hat{m}_{i'j'}}{\hat{m}_{ij'}\hat{m}_{i'j}}$$

をオッズ比とすることもある．

1.4 対数線形モデル

対数線型モデル（log-linear model）は，分割表の各セルの期待度数を対数変換した値を，各属性（因子）の**主効果**と属性の組合せの**交互作用**の線型和で表すモデルである．具体的に書くと，$I \times J$ 分割表の場合，

$$\log(m_{ij}) = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} \quad i = 1, \ldots, I, j = 1, \ldots, J$$

という対数線形モデルが得られる．各項はそれぞれ，

$$\begin{cases} u & : \text{切片} \\ u_{1(i)} & : \text{因子 1 の第 } i \text{ 水準の主効果} \\ u_{2(j)} & : \text{因子 2 の第 } j \text{ 水準の主効果} \\ u_{12(ij)} & : \text{2 因子水準 } (i,j) \text{ の交互作用（2 因子交互作用）} \end{cases}$$

とよばれる．2 元分割表では 2 因子交互作用までだが，より高次元の分割表を扱う場合には 3 因子交互作用，4 因子交互作用，とより大きな次数の交互作用が自然に考えられる．

定義式から読み取れるように，対数線形モデルは，セルの期待度数の対数値は，該当する各効果が影響を及ぼしあった結果であるということを仮定している．対数線形モデルを扱う利点は，各セルに対する主効果や交互作用効果を具体的に解析でき，解釈がしやすい点である．

ただし，いま挙げたモデルを定義式通りにみると，モデルのパラメータ（母数）の個数は，$1 + I + J + IJ$ 個であり，表のセルの数 IJ を超過している．そこで，各パラメータに

$$\sum_{i=1}^{I} u_{1(i)} = \sum_{j=1}^{J} u_{2(j)} = \sum_{i=1}^{I} u_{12(ij)} = \sum_{j=1}^{J} u_{12(ij)} = 0$$

という制約を付ければ，自由なパラメータ数は

$$1 + (I-1) + (J-1) + (I-1)(J-1) = IJ$$

となり，表のセル数に一致する．これは，このモデルがデータに完全に当てはまることに他ならない．このように，データを完全に説明するのに十分なパラメータが存在するモデルのことを**飽和モデル**（saturated model）という．飽和モデルのもとでは，$\hat{m}_{ij} = n_{ij}$ となる．

一般に，対数線形モデルのパラメータには制約が課せられる．いまのように,「添字に関して和をとったら 0 になる」という制約がその中でも自然なものである．

$I \times J$ 分割表の積多項サンプリング，多項サンプリング，それぞれの場合での対数線形モデルを考察する．

積多項サンプリング

行和固定の積多項表に制約がないとき，$m_{ij} = n_{ij}$ だった．対数変換すると $\log(m_{ij}) = \log(n_{ij})$ であり，$u = u_{1(i)} = u_{2(j)} = 0, u_{12(ij)} = \log(n_{ij})$ とおくことで，形式的ではあるが，飽和対数線形モデルの形 $\log(m_{ij}) = u + u_{1(i)} + u_{2(j)} + u_{12(ij)}$ が保たれていることがわかる．

命題 1.4.1. 行和固定の $I \times J$ 分割表の積多項サンプリングに関して，

$$\log(m_{ij}) = u + u_{1(i)} + u_{2(j)} \quad \text{for all } i,j$$
$$\iff \quad p_{1j} = \cdots = p_{Ij} \quad \text{for all } j$$

(証明)

($\Longrightarrow$)

任意の i, j に対して $m_{ij} = n_{i\cdot} p_{ij}$ である．また，仮定より任意の i, j に対して

$$\begin{aligned} m_{ij} &= \exp\{u + u_{1(i)} + u_{2(j)}\} \\ &= a a_{1(i)} a_{2(j)} \quad (e^u = a, e^u_{1(i)} = a_{1(i)}, e^u_{2(j)} = a_{2(j)}) \end{aligned}$$

である．ここで，

$$\begin{aligned} a a_{1(i)} a_{2(\cdot)} = m_{i\cdot} &= \sum_{j=1}^{J} m_{ij} \\ &= \sum_{j=1}^{J} n_{i\cdot} p_{ij} \\ &= n_{i\cdot} p_{i\cdot} \\ &= n_{i\cdot} \end{aligned}$$

が成り立つ．さらに $p_{ij} = m_{ij}/n_{i\cdot}$ であるから，

$$\begin{aligned} p_{ij} &= \frac{a a_{1(i)} a_{2(j)}}{n_{i\cdot}} \\ &= \frac{a a_{1(i)} a_{2(j)}}{a a_{1(i)} a_{2(\cdot)}} \\ &= \frac{a_{2(j)}}{a_{2(\cdot)}} \end{aligned}$$

が任意の i について成り立つ．よって，任意の j に対して $a_{2(j)}/a_{2(\cdot)} = p_{1j} = \cdots = p_{Ij}$ である．

($\Longleftarrow$)

任意の i, j に対して $m_{ij} = n_{i\cdot} p_{ij}$ である．また，仮定より任意の j に対して $\pi_j = p_{1j} = \cdots = p_{Ij}$ となる π_j がある．これより，

$$\begin{aligned} m_{ij} &= n_{i\cdot} \pi_j \\ \therefore \log(m_{ij}) &= \log(n_{i\cdot}) + \log(\pi_j) \end{aligned}$$

となる．ここで，$u = 0, u_{1(i)} = \log(n_{i\cdot}), u_{2(j)} = \log(\pi_j)$ とおくことで $\log(m_{ij}) = u + u_{1(i)} + u_{2(j)}$ の形式の対数線形モデルが得られる． □

この命題により，積多項表の仮説 (1.1) の検定は，

$$\begin{cases} \mathrm{H}_0 : \quad \log(m_{ij}) = u + u_{1(i)} + u_{2(j)} \quad \text{for all } i, j \\ \mathrm{H}_1 : \quad \log(m_{ij}) = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} \quad \text{for all } i, j \end{cases}$$

という対数線形モデルを対象にした形式，もしくはよりシンプルに

$$\begin{cases} \mathrm{H}_0 : \quad u_{12(ij)} = 0 \text{ for all } i, j \\ \mathrm{H}_1 : \quad \mathrm{H}_0 \text{は偽} \end{cases}$$

の検定とも書けることがわかる．

多項サンプリング

積多項表と同様，制約がないときは $m_{ij} = n_{ij}$ であるから，形式的ではあるが，飽和対数線形モデルの形 $\log(m_{ij}) = u + u_{1(i)} + u_{2(j)} + u_{12(ij)}$ が保たれている．

命題 1.4.2. $I \times J$ 分割表の多項サンプリングに関して，

$$\begin{aligned} &\log(m_{ij}) = u + u_{1(i)} + u_{2(j)} \quad \text{for all } i, j \\ \iff \quad &p_{ij} = p_{i\cdot} p_{\cdot j} \quad \text{for all } i, j \end{aligned}$$

(証明)
($\Longrightarrow$)
仮定より，

$$\begin{aligned} m_{ij} &= e^{u + u_{1(i)} + u_{2(j)}} \\ &= a a_{1(i)} a_{2(j)} \qquad (a = e^u, a_{1(i)} = e^{u_{1(i)}}, a_{2(j)} = e^{u_{2(j)}}) \end{aligned}$$

である．すると $p_{ij} = \dfrac{m_{ij}}{n_{\cdot\cdot}} = \dfrac{a a_{1(i)} a_{2(j)}}{n_{\cdot\cdot}}$ である．これより，

$$p_{i\cdot} = \frac{a a_{1(i)} a_{2(\cdot)}}{n_{\cdot\cdot}}, \qquad p_{\cdot j} = \frac{a a_{1(\cdot)} a_{2(j)}}{n_{\cdot\cdot}}$$

となる．また同様に $1 = p_{\cdot\cdot} = \dfrac{a a_{1(\cdot)} a_{2(\cdot)}}{n_{\cdot\cdot}}$ となる．
以上より，

$$\begin{aligned} p_{i\cdot} p_{\cdot j} &= \frac{a a_{1(i)} a_{2(\cdot)} a a_{1(\cdot)} a_{2(j)}}{n_{\cdot\cdot}^2} \\ &= \left\{ \frac{a a_{1(i)} a_{2(j)}}{n_{\cdot\cdot}} \right\} \left\{ \frac{a a_{2(\cdot)} a_{1(\cdot)}}{n_{\cdot\cdot}} \right\} \\ &= p_{ij} p_{\cdot\cdot} \\ &= p_{ij} \end{aligned}$$

となる．
($\Longleftarrow$)
任意の i, j に対して $m_{ij} = n_{\cdot\cdot} p_{ij}$ である．これと仮定より，$m_{ij} = n_{\cdot\cdot} p_{i\cdot} p_{\cdot j}$ であるから，

$$\log(m_{ij}) = \log(n_{\cdot\cdot}) + \log(p_{i\cdot}) + \log(p_{\cdot j})$$

となり，$\log(m_{ij}) = u + u_{1(i)} + u_{2(j)}$ の形式の対数線形モデルとなる． □
この命題により，積多項表と同様，多項表の仮説 (1.2) の検定は，

$$\begin{cases} \mathrm{H}_0 : & \log(m_{ij}) = u + u_{1(i)} + u_{2(j)} \quad \text{for all } i, j \\ \mathrm{H}_1 : & \log(m_{ij}) = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} \quad \text{for all } i, j \end{cases}$$

もしくは,

$$\begin{cases} \mathrm{H}_0: & u_{12(ij)} = 0 \quad \text{for all } i, j \\ \mathrm{H}_1: & \mathrm{H}_0\text{は偽} \end{cases}$$

の検定とも書けることがわかる.

1.5 データ解析例

統計解析向けプログラミング言語 R を用いて，実際のデータに対数線形モデルを当てはめてみる.

次のデータは，2003 年 11〜12 月に首都圏の中学 1 〜 3 年生を対象に行った，学校の授業中に関するアンケートの結果である．こちらは，資料 [8]（CRN のホームページ）より引用している．なお，もとのデータがパーセンテージで与えられていたため，今回は解析のために度数に変換している.

	Q 1. 授業中，居眠りをすることがあるか				
	1. よくする	2. ときどきする	3. あまりしない	4. ぜったいしない	計
男子	38	184	224	371	817
女子	26	164	241	313	744

	Q 2. 授業中，手紙を回すことがあるか				
	1. よくする	2. ときどきする	3. あまりしない	4. ぜったいしない	計
男子	9	25	66	717	817
女子	45	146	212	341	744

どちらの表も，属性が性別・回答の 2 つで，行和が固定の 2 元積多項サンプリング表と見なせる．また属性については，属性 1(i):性別（1=男子，2=女子），属性 2(j):回答（1=よくする，$\cdots$，4=ぜったいしない）としておく.

以下は，R による 2 つの対数線形モデル

$$\log(m_{ij}) = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} \qquad \text{(モデル 1（飽和モデル))}$$

$$\log(m_{ij}) = u + u_{1(i)} + u_{2(j)} \qquad \text{(モデル 2)}$$

の適合の結果である．なお，以下の school-Q1(Q2).csv ファイルとは，上の表データを R での解析用に編集したものである.

まず，質問 1 の表の解析を行う．こちらは，R の関数 loglm を利用する．この関数は，分割表データに設定した対数線形モデルを当てはめ，当てはめ値（そのモデルでの期待度数）や，検定統計量等を算出する.

まずモデル 1 を当てはめる.

```
> Q1 <- read.csv("school-Q1.csv",header=T)
        #質問１の表データを読み込み,「Q1」と名前を付ける
> Q1
   Sex Answer   N   #元の表と形式は異なるが，同じ意味である
1  Boy      A  38   #関数 loglm を利用するには，この形式にする必要がある
2  Boy      B 184
3  Boy      C 224
4  Boy      D 371
5 Girl      A  26
6 Girl      B 164
7 Girl      C 241
8 Girl      D 313
>
>
> Q1.m1 <- loglm(N~Sex+Answer+Sex*Answer,data=Q1)
           #モデル１（飽和モデル）を当てはめる．Q1.m1 と名を付けた
> Q1.m1   #呼び出し
Call:
loglm(formula = N ~ Sex + Answer + Sex * Answer, data = Q1)

Statistics:
                 X^2 df P(> X^2)
Likelihood Ratio   0  0        1   #尤度比検定統計量
Pearson            0  0        1   #ピアソンカイ二乗統計量
> fitted(Q1.m1)   #当てはめ値
Re-fitting to get fitted values
      Answer
Sex     A   B   C   D
  Boy  38 184 224 371   #元のデータと一致している
  Girl 26 164 241 313
> residuals(Q1.m1)   #残差
Re-fitting to get frequencies and fitted values
      Answer
Sex    A B C D
  Boy  0 0 0 0
```

```
  Girl 0 0 0 0
```

loglm(N Sex+Answer+Sex*Answer,data=Q1) は,「Q1」データに，性別（Sex）の主効果 $u_1(i)$，回答（Answer）の主効果 $u_2(j)$，性別と回答の交互作用 $u_{12(ij)}$ が含まれる対数線形モデルを当てはめるという意味である．切片項 u は自然に含まれている．
飽和モデルであるから，完全に当てはまっている．これは，検定統計量の P 値，当てはめ値，残差の結果から伺える．

次にモデル２を当てはめる．

```
> Q1.m2 <- loglm(N~Sex+Answer,data=Q1)   #モデル２
> Q1.m2
Call:
loglm(formula = N ~ Sex + Answer, data = Q1)

Statistics:
                      X^2 df  P(> X^2)
Likelihood Ratio 5.544032  3 0.1360286
Pearson          5.537332  3 0.1364227
> fitted(Q1.m2)
Re-fitting to get fitted values
      Answer
Sex           A        B        C        D
  Boy  33.49648 182.1371 243.3728 357.9936
  Girl 30.50352 165.8629 221.6272 326.0064
> residuals(Q1.m2)
Re-fitting to get frequencies and fitted values
      Answer
Sex             A          B         C          D
  Boy   0.7616047  0.1378017 -1.258863  0.6833154
  Girl -0.8368228 -0.1449215  1.283013 -0.7252224
```

検定統計量の P 値は，約 0.136 ほどである．これより，仮説 $\mathrm{H}_0 : u_{12(ij)} = 0$ は有意でなく，つまりモデル２の当てはまりが悪くないことがわかる．

この結果から,「授業中の居眠り」に関して，首都圏中学生の男女間では統計的に意味のある差は認められないと判断できる．

次に質問２の表の解析を行う．こちらは，R の関数 glm を利用する．この関数は，一般化線形モデルをデータに当てはめる関数である．以下では，関数の引数で対数線形モデルを当てはめる

ように設定している．また，この関数は尤度比検定統計量・当てはめ値のほか，モデルパラメータの推定値も算出する．

まず，モデル１を当てはめる．

```
> Q2 <- read.csv("school-Q2.csv",header=T)  #質問２データの読み込み
> Q2
   Sex Answer   N
1  Boy      A   9
2  Boy      B  25
3  Boy      C  66
4  Boy      D 717
5 Girl      A  45
6 Girl      B 146
7 Girl      C 212
8 Girl      D 341
>
>
> Q2.m1 <- glm(N~Sex+Answer+Sex*Answer,family=poisson,
               contrasts=list(Sex="contr.sum",Answer="contr.sum"),data=Q2)
               #family=poisson は対数線形を当てはめるという指定.
               #モデル１
> summary(Q2.m1) #結果呼び出し

Call:
glm(formula = N ~ Sex + Answer + Sex * Answer, family = poisson,
    data = Q2, contrasts = list(Sex = "contr.sum", Answer = "contr.sum"))

Deviance Residuals:
[1]  0  0  0  0  0  0  0  0

Coefficients:   #各パラメータ推定値
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   4.51995    0.05651  79.984  < 2e-16 ***
Sex1         -0.47474    0.05651  -8.401  < 2e-16 ***
Answer1      -1.51800    0.14093 -10.772  < 2e-16 ***
Answer2      -0.41870    0.09513  -4.401 1.08e-05 ***
```

```
Answer3       0.25317    0.07535   3.360 0.000779 ***
Sex1:Answer1 -0.32998    0.14093  -2.342 0.019205 *
Sex1:Answer2 -0.40763    0.09513  -4.285 1.83e-05 ***
Sex1:Answer3 -0.10873    0.07535  -1.443 0.149007
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance:  1.7642e+03  on 7  degrees of freedom
Residual deviance: -1.1813e-13  on 0  degrees of freedom #尤度比検定統計量
AIC: 66.897

Number of Fisher Scoring iterations: 3

> fitted(Q2.m1)
  1   2   3   4   5   6   7   8
  9  25  66 717  45 146 212 341
```

glm 関数内の contrasts=list(Sex="contr.sum",Answer="contr.sum") は，主効果と交互作用に「添字について和をとったら 0 になる」という制約を付けるという指定である．summary 関数では，制約により自然に決まるパラメータについては表示されないので，それらも手計算し，合わせてパラメータの推定結果として次の表にまとめた．
「R での表示」の欄の"-"は，summary 関数で表示されていないという意味である．

効果項	R での表示	パラメータ	推定値
切片	Intercept	u	4.51995
第 1 主効果	Sex1	$u_{1(1)}$	-0.47474
	-	$u_{1(2)}$	0.47474
第 2 主効果	Answer1	$u_{2(1)}$	-1.518
	Answer2	$u_{2(2)}$	-0.4187
	Answer3	$u_{2(3)}$	0.25317
	-	$u_{2(4)}$	1.68353
交互作用	Sex1:Answer1	$u_{12(11)}$	-0.32998
	Sex1:Answer2	$u_{12(12)}$	-0.40763
	Sex1:Answer3	$u_{12(13)}$	-0.10873
	-	$u_{12(14)}$	0.84634
	-	$u_{12(21)}$	0.32998
	-	$u_{12(22)}$	0.40763
	-	$u_{12(23)}$	0.10873
	-	$u_{12(24)}$	-0.84634

これより，例えばセル (1,1) の期待度数は,

$$\begin{aligned}\log(m_{11}) &= u + u_{1(1)} + u_{2(1)} + u_{12(11)} \\ &= \{(4.51995) + (-0.47474) + (-1.518) + (-0.32998)\}\end{aligned}$$

と書け，実際に計算すると,

$$m_{11} = 9.000048804$$

となる．これは，セル (1,1) の度数（観測値）とほぼ一致しており，飽和モデルが，完全に当てはまることが確認できる.

次にモデル 2 を当てはめる.

```
> Q2.m2 <- glm(N~Sex+Answer,family="poisson",
               contrasts=list(Sex="contr.sum",Answer="contr.sum"),data=Q2)
> summary(Q2.m2)

Call:
glm(formula = N ~ Sex + Answer, family = "poisson", data = Q2,
    contrasts = list(Sex = "contr.sum", Answer = "contr.sum"))

Deviance Residuals:
     1       2       3       4       5       6       7       8
-4.234  -8.076  -7.393   6.633   3.429   6.421   6.347  -7.728
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.73636    0.04252 111.384  < 2e-16 ***
Sex1         0.04680    0.02534   1.847  0.06475 .
Answer1     -1.44162    0.10520 -13.704  < 2e-16 ***
Answer2     -0.28894    0.06878  -4.201 2.66e-05 ***
Answer3      0.19702    0.06004   3.281  0.00103 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1764.2  on 7  degrees of freedom
Residual deviance:  334.8  on 3  degrees of freedom #尤度比検定統計量
AIC: 395.70

Number of Fisher Scoring iterations: 5

> fitted(Q2.m2)
        1         2         3         4         5         6         7         8
 28.26265  89.49840 145.50032 553.73863  25.73735  81.50160 132.49968 504.26137
```

尤度比検定統計量は，自由度 3 で 334.8 と算出された．このときの P 値は，R により

```
> 1-pchisq(334.8,3)
[1] 0
```

と算出される．pchisq(334.8,3) とは，自由度 3 のカイ二乗分布における点 334.8 の下側確率を算出する命令である．またこの分布の上側 5% 点は，R により

```
> qchisq(0.95,3)
[1] 7.814728
```

と算出される．これからもモデル 2 の当てはまりの悪さがわかる．すなわち $u_{12(ij)} = 0$ とおくことはできないといえる．

解析の結果，質問 2 の表に対しては，飽和モデルが選択される，すなわち,「授業中に手紙をまわす」という行為には男女間に差があるということが判断される．交互作用に注目すると，

$$\begin{array}{ll} \text{（男子）} & u_{12(11)} = -0.33, \quad u_{12(12)} = -0.41 \\ \text{（女子）} & u_{12(21)} = 0.33, \quad\ \ u_{12(22)} = 0.41 \end{array}$$

となっている．効果は，男子は負，女子は正であることから，女子は手紙を回す人数が多くなる傾向があると解釈できる．

今回の例の場合，質問２の方は自身の経験からも女子の方が多くなるという予想はしやすいものであったが，質問１の居眠りに差はないという結果は，予想に反していた．

データに対して予備知識がなくとも，分割表解析によりある程度の統計的な考察が可能である．また，R により正確な解析が可能である．

第2章　3元分割表

2元分割表と同様，3つの属性で分類した分割表を3元分割表とよぶ．本章では，対象を $I \times J \times K$ 分割表とし，3つの属性を A, B, C とする．また，各属性の水準を示す添字として順に i, j, k を割り当てることにする．3元分割表のセル (i, j, k) の度数は，$n_{ijk},\ i = 1, \ldots, I,\ j = 1 \ldots, J,\ k = 1, \ldots, K$ で表わす．3元分割表は2元分割表を一段階拡張した形式で表わせる．下の表は度数 n_{ijk} の $I \times J \times K$ 分割表の形式である．

		属性 C						
		C_1			$\ldots$	C_K		
	属性 B	B_1	$\cdots$	B_J	$\ldots$	B_1	$\cdots$	B_J
属性 A	A_1	n_{111}	$\cdots$	n_{1J1}		n_{11K}	$\cdots$	n_{1JK}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\cdots$	$\vdots$	$\ddots$	$\vdots$
	A_I	n_{I11}	$\cdots$	n_{IJ1}		n_{I1K}	$\cdots$	n_{IJK}

周辺和，総和は次のように表わされる．

$$
\begin{cases}
n_{ij\cdot} = \displaystyle\sum_{k=1}^{K} n_{ijk},\ n_{i\cdot k} = \sum_{j=1}^{J} n_{ijk},\ n_{\cdot jk} = \sum_{i=1}^{I} n_{ijk} \\
n_{i\cdot\cdot} = \displaystyle\sum_{j=1}^{J}\sum_{k=1}^{K} n_{ijk},\ n_{\cdot j\cdot} = \sum_{i=1}^{I}\sum_{k=1}^{K} n_{ijk},\ n_{\cdot\cdot k} = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ijk} \\
n_{\cdots} = \displaystyle\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} n_{ijk}
\end{cases}
$$

2元分割表と同様，確率 p_{ijk}, 期待度数 m_{ijk} についても上と同様の表がある．

本章では，この表をもとに3元分割表解析について考察する．ただし，基本的に考察の対象を多項サンプリング表（総和が固定）に限定する．

周辺確率

3元分割表のモデルを考える前に，周辺確率の MLE を与える．次は，k について和をとった $n_{ij\cdot}$ を並べた $I \times J$ 表である．

		属性 $C = C_\cdot$			
	属性 B	B_1	$\cdots$	B_J	計
属性 A	A_1	$n_{11\cdot}$	$\cdots$	$n_{1J\cdot}$	$n_{1\cdot\cdot}$
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	A_I	$n_{I1\cdot}$	$\cdots$	$n_{IJ\cdot}$	$n_{I\cdot\cdot}$
	計	$n_{\cdot 1\cdot}$	$\cdots$	$n_{\cdot J\cdot}$	$n_{\cdots}$

同様の表が，p についても得られる．ここで，第 1.2 節に記述した $\boldsymbol{p}$ の MLE の算出法を用いると，この表の確率・周辺確率の MLE は次のように導かれる．

$$\hat{p}_{ij\cdot} = \frac{n_{ij\cdot}}{n_{\cdots}}$$
$$\hat{p}_{i\cdot\cdot} = \frac{n_{i\cdot\cdot}}{n_{\cdots}}\ ,\ \hat{p}_{\cdot j\cdot} = \frac{n_{\cdot j\cdot}}{n_{\cdots}}$$

上では，属性 C をまとめて $I \times J$ 表にしたが，同様の操作は A, B に対しても行うことができる．これより，$I \times J \times K$ 分割表において，とりうる全ての周辺確率の MLE が導かれる．結果をまとめると

$$\begin{cases} \hat{p}_{i\cdot\cdot} = \dfrac{n_{i\cdot\cdot}}{n_{\cdots}}\ ,\ \hat{p}_{\cdot j\cdot} = \dfrac{n_{\cdot j\cdot}}{n_{\cdots}}\ ,\ \hat{p}_{\cdot\cdot k} = \dfrac{n_{\cdot\cdot k}}{n_{\cdots}}\ , \\ \hat{p}_{ij\cdot} = \dfrac{n_{ij\cdot}}{n_{\cdots}}\ ,\ \hat{p}_{i\cdot k} = \dfrac{n_{i\cdot k}}{n_{\cdots}}\ ,\ \hat{p}_{\cdot jk} = \dfrac{n_{\cdot jk}}{n_{\cdots}} \end{cases} \quad i = 1, \ldots, I,\ j = 1, \ldots, J,\ k = 1, \ldots, K$$

となる．これら周辺確率の MLE は，３元分割表のモデルのセル確率 p_{ijk} の MLE の算出に利用する．

2.1 ３元分割表のモデル

３元分割表にもいくつか考えられうる統計モデルが存在する．これを以下に列挙する．

$M^{(1)}$: 完全独立モデル（A, B, C が互いに独立なモデル）

$M^{(2)}$: 周辺独立モデル 1（A と (B, C) が独立なモデル）

$M^{(3)}$: 周辺独立モデル 2（B と (A, C) が独立なモデル）

$M^{(4)}$: 周辺独立モデル 3（C と (A, B) が独立なモデル）

$M^{(5)}$: 条件付独立モデル 1（C を与えたもとで A, B が条件付独立であるモデル）

$M^{(6)}$: 条件付独立モデル 2（B を与えたもとで A, C が条件付独立であるモデル）

$M^{(7)}$: 条件付独立モデル 3（A を与えたもとで B, C が条件付独立であるモデル）

$M^{(8)}$: オッズ比均一モデル

2.1.1 完全独立モデル

完全独立モデルは，その名の通り A, B, C が互いに独立（$A \perp\!\!\!\perp B \perp\!\!\!\perp C$）であるモデルのことである．具体的に書くと，３元分割表の任意のセル (i, j, k) の確率 p_{ijk} が，

$$M^{(1)} : p_{ijk} = p_{i\cdot\cdot} p_{\cdot j\cdot} p_{\cdot\cdot k}$$

となるモデル $M^{(1)}$ を完全独立モデルという．$M^{(1)}$ のもとでの p_{ijk} の最尤推定値 $\hat{p}_{ijk}^{(1)}$ は，

$$\begin{aligned} \hat{p}_{ijk}^{(1)} &= \hat{p}_{i\cdot\cdot} \hat{p}_{\cdot j\cdot} \hat{p}_{\cdot\cdot k} \\ &= \frac{n_{i\cdot\cdot}}{n_{\cdots}} \cdot \frac{n_{\cdot j\cdot}}{n_{\cdots}} \cdot \frac{n_{\cdot\cdot k}}{n_{\cdots}} \end{aligned}$$

となる．またこれより，$M^{(1)}$ のもとでの m_{ijk} の最尤推定値 $\hat{m}_{ijk}^{(1)}$ は，

$$\begin{aligned} \hat{m}_{ijk}^{(1)} &= n_{\cdots} \hat{p}_{ijk}^{(1)} \\ &= \frac{n_{i\cdot\cdot} n_{\cdot j\cdot} n_{\cdot\cdot k}}{n_{\cdots}^2} \end{aligned}$$

となる．このことから，モデル $M^{(1)}$ には，

$$\hat{m}_{i\cdot\cdot} = n_{i\cdot\cdot},\ \hat{m}_{\cdot j\cdot} = n_{\cdot j\cdot},\ \hat{m}_{\cdot\cdot k} = n_{\cdot\cdot k}$$

という期待度数に関する周辺制約が存在することがわかる．

2.1.2 周辺独立モデル

周辺独立モデルは，３つの属性のうち１つの属性が残り２つの属性と独立なモデルである．周辺独立モデルの３つのパターンはそれぞれ具体的に書くと，

$$\begin{aligned} M^{(2)} &: p_{ijk} = p_{i\cdot\cdot} p_{\cdot jk}, \quad A \perp\!\!\!\perp (B, C) \\ M^{(3)} &: p_{ijk} = p_{\cdot j\cdot} p_{i\cdot k}, \quad B \perp\!\!\!\perp (A, C) \\ M^{(4)} &: p_{ijk} = p_{\cdot\cdot k} p_{ij\cdot}, \quad C \perp\!\!\!\perp (A, B) \end{aligned}$$

となる．
$M^{(2)}$ のもとでの p_{ijk} の最尤推定値 $\hat{p}_{ijk}^{(2)}$ は，

$$\begin{aligned} \hat{p}_{ijk}^{(2)} &= \hat{p}_{i\cdot\cdot} \hat{p}_{\cdot jk} \\ &= \frac{n_{i\cdot\cdot}}{n_{\cdots}} \cdot \frac{n_{\cdot jk}}{n_{\cdots}} \end{aligned}$$

となる．またこれより，$M^{(2)}$ のもとでの m_{ijk} の最尤推定値 $\hat{m}_{ijk}^{(2)}$ は，

$$\begin{aligned} \hat{m}_{ijk}^{(2)} &= n_{\cdots} \hat{p}_{ijk}^{(2)} \\ &= \frac{n_{i\cdot\cdot} n_{\cdot jk}}{n_{\cdots}} \end{aligned}$$

となる．このことから，モデル $M^{(2)}$ には，

$$\hat{m}_{i\cdot\cdot} = n_{i\cdot\cdot},\ \hat{m}_{\cdot jk} = n_{\cdot jk}$$

という期待度数に関する周辺制約が存在することがわかる．
$M^{(3)}$,$M^{(4)}$ についても同様の議論が行われる．最尤推定値と制約について，結果をまとめる．

$$\begin{aligned}
M^{(2)}:&\quad \hat{p}^{(2)}_{ijk} = \frac{n_{i\cdot\cdot}n_{\cdot jk}}{n^2_{\cdots}},\quad \hat{m}^{(2)}_{ijk} = \frac{n_{i\cdot\cdot}n_{\cdot jk}}{n_{\cdots}} \quad (\hat{m}_{i\cdot\cdot} = n_{i\cdot\cdot}, \hat{m}_{\cdot jk} = n_{\cdot jk})\\
M^{(3)}:&\quad \hat{p}^{(3)}_{ijk} = \frac{n_{\cdot j\cdot}n_{i\cdot k}}{n^2_{\cdots}},\quad \hat{m}^{(3)}_{ijk} = \frac{n_{\cdot j\cdot}n_{i\cdot k}}{n_{\cdots}} \quad (\hat{m}_{\cdot j\cdot} = n_{\cdot j\cdot}, \hat{m}_{i\cdot k} = n_{i\cdot k})\\
M^{(4)}:&\quad \hat{p}^{(4)}_{ijk} = \frac{n_{\cdot\cdot k}n_{ij\cdot}}{n^2_{\cdots}},\quad \hat{m}^{(4)}_{ijk} = \frac{n_{\cdot\cdot k}n_{ij\cdot}}{n_{\cdots}} \quad (\hat{m}_{\cdot\cdot k} = n_{\cdot\cdot k}, \hat{m}_{ij\cdot} = n_{ij\cdot})
\end{aligned}$$

2.1.3 条件付独立モデル

条件付独立モデルは，１つの属性を与えたもとで残り２つ属性が条件付独立となるモデルである．C を与えたもとで A, B が条件付独立である場合（$A \perp\!\!\!\perp B \mid C$）のセル (i, j, k) の確率を考える．条件付確率の定義より，$A \perp\!\!\!\perp B \mid C$ のとき

$$\Pr(A = A_i, B = B_j \mid C = C_k) = \Pr(A = A_i \mid C = C_k)\Pr(B = B_j \mid C = C_k)$$

であるから，

$$\frac{\Pr(A = A_i, B = B_j, C = C_k)}{\Pr(C = C_k)} = \frac{\Pr(A = A_i, C = C_k)}{\Pr(C = C_k)}\frac{\Pr(B = B_j, C = C_k)}{\Pr(C = C_k)}$$

$$\therefore \Pr(A = A_i, B = B_j, C = C_k) = \frac{\Pr(A = A_i, C = C_k)\cdot\Pr(B = B_j, C = C_k)}{\Pr(C = C_k)}$$

となる．よって $A \perp\!\!\!\perp B \mid C$ のときセル (i, j, k) の確率は，

$$p_{ijk} = \frac{p_{i\cdot k}p_{\cdot jk}}{p_{\cdot\cdot k}}$$

となる．$A \perp\!\!\!\perp C \mid B, B \perp\!\!\!\perp C \mid A$ の場合も同様の手順で p_{ijk} が求められる．
条件付独立モデルの３つのパターンはそれぞれ具体的に書くと，

$$\begin{aligned}
M^{(5)}: p_{ijk} &= \frac{p_{i\cdot k}p_{\cdot jk}}{p_{\cdot\cdot k}},\quad A \perp\!\!\!\perp B \mid C\\
M^{(6)}: p_{ijk} &= \frac{p_{ij\cdot}p_{\cdot jk}}{p_{\cdot j\cdot}},\quad A \perp\!\!\!\perp C \mid B\\
M^{(7)}: p_{ijk} &= \frac{p_{ij\cdot}p_{i\cdot k}}{p_{i\cdot\cdot}},\quad B \perp\!\!\!\perp C \mid A
\end{aligned}$$

となる．
$M^{(5)}$ のもとでの p_{ijk} の最尤推定値 $\hat{p}^{(5)}_{ijk}$ は，

$$\hat{p}^{(5)}_{ijk} = \frac{\hat{p}_{i\cdot k}\hat{p}_{\cdot jk}}{\hat{p}_{\cdot\cdot k}}$$

$$= \left(\frac{n_{i\cdot k}}{n_{\cdots}} \cdot \frac{n_{\cdot jk}}{n_{\cdots}}\right) \Big/ \left(\frac{n_{\cdot\cdot k}}{n_{\cdots}}\right)$$
$$= \frac{n_{i\cdot k} n_{\cdot jk}}{n_{\cdot\cdot k} n_{\cdots}}$$

となる．これより $M^{(5)}$ のもとでの m_{ijk} の最尤推定値 $\hat{m}^{(5)}_{ijk}$ は，

$$\hat{m}^{(5)}_{ijk} = n_{\cdots} \cdot \frac{\hat{p}_{i\cdot k}\hat{p}_{\cdot jk}}{\hat{p}_{\cdot\cdot k}}$$
$$= \frac{n_{i\cdot k} n_{\cdot jk}}{n_{\cdot\cdot k}}$$

となる．またこのことから，モデル $M^{(5)}$ には，

$$\hat{m}_{i\cdot k} = n_{i\cdot k},\ \hat{m}_{\cdot jk} = n_{\cdot jk}$$

という期待度数に関する周辺制約が存在することがわかる．
$M^{(6)}$,$M^{(7)}$ についても同様の議論が行われる．最尤推定値と制約について，結果をまとめる．

$$M^{(5)}:\ \hat{p}^{(5)}_{ijk} = \frac{n_{i\cdot k} n_{\cdot jk}}{n_{\cdot\cdot k} n_{\cdots}},\quad \hat{m}^{(5)}_{ijk} = \frac{n_{i\cdot k} n_{\cdot jk}}{n_{\cdot\cdot k}} \quad (\hat{m}_{i\cdot k} = n_{i\cdot k},\ \hat{m}_{\cdot jk} = n_{\cdot jk})$$
$$M^{(6)}:\ \hat{p}^{(6)}_{ijk} = \frac{n_{ij\cdot} n_{\cdot jk}}{n_{\cdot j\cdot} n_{\cdots}},\quad \hat{m}^{(6)}_{ijk} = \frac{n_{ij\cdot} n_{\cdot jk}}{n_{\cdot j\cdot}} \quad (\hat{m}_{ij\cdot} = n_{ij\cdot},\ \hat{m}_{\cdot jk} = n_{\cdot jk})$$
$$M^{(7)}:\ \hat{p}^{(7)}_{ijk} = \frac{n_{ij\cdot} n_{i\cdot k}}{n_{i\cdot\cdot} n_{\cdots}},\quad \hat{m}^{(7)}_{ijk} = \frac{n_{ij\cdot} n_{i\cdot k}}{n_{i\cdot\cdot}} \quad (\hat{m}_{ij\cdot} = n_{ij\cdot},\ \hat{m}_{i\cdot k} = n_{i\cdot k})$$

2.1.4 オッズ比均一モデル

3元分割表で考えられる興味のあるモデルで，オッズ比によって表現されるものがある．具体的に書くと，

$$M^{(8)}: \frac{p_{111}p_{ij1}}{p_{i11}p_{1j1}} = \frac{p_{11k}p_{ijk}}{p_{i1k}p_{1jk}} \quad \text{for all } i = 2,\cdots,I,\ j = 2,\cdots,J,\ k = 2,\cdots,K$$

を満たすモデルである．このモデルのことを**オッズ比均一モデル**とよぶことにする．
$M^{(8)}$ のもとでの m_{ijk} の最尤推定値 $\hat{m}^{(8)}_{ijk}$ には，周辺制約

$$\hat{m}^{(8)}_{ij\cdot} = n_{ij\cdot},\ \hat{m}^{(8)}_{i\cdot k} = n_{i\cdot k},\ \hat{m}^{(8)}_{\cdot jk} = n_{\cdot jk}$$

が存在する．この制約の存在を，3元分割表の最小のサイズである $2\times2\times2$ 表の場合で確かめる．

	k	1		2	
	j	1	2	1	2
i	1	n_{111}	n_{121}	n_{112}	n_{122}
	2	n_{211}	n_{221}	n_{212}	n_{222}

命題 2.1.1. $2\times2\times2$ 分割表における $M^{(8)}$ のもとでの m_{ijk} の最尤推定値 $\hat{m}_{ijk}$ には，周辺制約

$$\hat{m}_{ij\cdot}=n_{ij\cdot},\ \hat{m}_{i\cdot k}=n_{i\cdot k},\ \hat{m}_{\cdot jk}=n_{\cdot jk}$$

が存在する．

（証明）
$2\times2\times2$ 分割表のもとでは, $M^{(8)}$ の p_{ijk} の条件式は,

$$\frac{p_{111}p_{221}}{p_{211}p_{121}}=\frac{p_{112}p_{222}}{p_{212}p_{122}}$$

と書ける．この式より,

$$p_{111}p_{221}p_{212}p_{122}-p_{112}p_{222}p_{211}p_{121}=0 \tag{2.1}$$

と書ける．

また，多項サンプリング表の性質 $\sum_{i,j,k}p_{ijk}=1$ より,

$$\sum_i\sum_j\sum_k p_{ijk}-1=0 \tag{2.2}$$

と書ける．

確率 $\boldsymbol{p}$ の MLE は尤度関数 $L(\boldsymbol{p})$ を最大にする点 $\hat{\boldsymbol{p}}$ であった．２元分割表のときと同様，３元多項サンプリング表での $\boldsymbol{p}$ の対数尤度関数 $\log L(\boldsymbol{p})$ の $\boldsymbol{p}$ に依存する項は,

$$\ell(\boldsymbol{p})=\sum_i\sum_j\sum_k n_{ijk}\log p_{ijk}$$

である．

条件 (2.1)，(2.2) のもとで $\ell(\boldsymbol{p})$ を最大にする $\hat{\boldsymbol{p}}$ を求めるために，ラグランジュの未定乗数法を用いる．関数 F を,

$$F(\boldsymbol{p},\lambda_1,\lambda_2)=\sum_{ijk}n_{ijk}\log p_{ijk}-\lambda_1\left(p_{111}p_{221}p_{212}p_{122}-p_{112}p_{222}p_{211}p_{121}\right)-\lambda_2\left(\sum_{ijk}p_{ijk}-1\right)$$

とおく．この F を $\boldsymbol{p}$ の要素と λ_1,λ_2 でそれぞれ偏微分して 0 とおくと,

$$\begin{aligned}
\frac{\partial F}{\partial p_{111}}&=\frac{n_{111}}{p_{111}}-\lambda_1p_{221}p_{212}p_{122}-\lambda_2=0\\
\frac{\partial F}{\partial p_{221}}&=\frac{n_{221}}{p_{221}}-\lambda_1p_{111}p_{212}p_{122}-\lambda_2=0\\
\frac{\partial F}{\partial p_{212}}&=\frac{n_{212}}{p_{212}}-\lambda_1p_{111}p_{221}p_{122}-\lambda_2=0\\
\frac{\partial F}{\partial p_{122}}&=\frac{n_{122}}{p_{122}}-\lambda_1p_{111}p_{221}p_{212}-\lambda_2=0\\
\frac{\partial F}{\partial p_{112}}&=\frac{n_{112}}{p_{112}}+\lambda_1p_{222}p_{211}p_{121}-\lambda_2=0
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial F}{\partial p_{222}} &= \frac{n_{222}}{p_{222}} + \lambda_1 p_{112}p_{211}p_{121} - \lambda_2 = 0 \\
\frac{\partial F}{\partial p_{211}} &= \frac{n_{211}}{p_{211}} + \lambda_1 p_{112}p_{222}p_{121} - \lambda_2 = 0 \\
\frac{\partial F}{\partial p_{121}} &= \frac{n_{121}}{p_{121}} + \lambda_1 p_{112}p_{222}p_{211} - \lambda_2 = 0 \\
\frac{\partial F}{\partial \lambda_1} &= -p_{111}p_{221}p_{212}p_{122} + p_{112}p_{222}p_{211}p_{121} = 0 \\
\frac{\partial F}{\partial \lambda_2} &= -\sum_{i,j,k} p_{ijk} + 1 = 0
\end{aligned}
$$

となる．ラグランジュの未定乗数法より，この連立方程式の解 $\boldsymbol{p}$=$\hat{\boldsymbol{p}}$ が制約 (2.1),(2.2) のもとで $\ell(\boldsymbol{p})$ を最大にする点，つまりモデル $M^{(8)}$ の確率の MLE である．

8 つの式 $\dfrac{\partial F}{\partial p_{ijk}} = 0$ にそれぞれ p_{ijk} を掛けると，

$$
\begin{cases}
n_{111} - \lambda_1 p_{111}p_{221}p_{212}p_{122} - \lambda_2 p_{111} = 0 \\
n_{221} - \lambda_1 p_{111}p_{221}p_{212}p_{122} - \lambda_2 p_{221} = 0 \\
n_{212} - \lambda_1 p_{111}p_{221}p_{212}p_{122} - \lambda_2 p_{212} = 0 \\
n_{122} - \lambda_1 p_{111}p_{221}p_{212}p_{122} - \lambda_2 p_{122} = 0 \\
n_{112} + \lambda_1 p_{112}p_{222}p_{211}p_{121} - \lambda_2 p_{112} = 0 \\
n_{222} + \lambda_1 p_{112}p_{222}p_{211}p_{121} - \lambda_2 p_{222} = 0 \\
n_{211} + \lambda_1 p_{112}p_{222}p_{211}p_{121} - \lambda_2 p_{211} = 0 \\
n_{121} + \lambda_1 p_{112}p_{222}p_{211}p_{121} - \lambda_2 p_{121} = 0
\end{cases}
\tag{2.3}
$$

となる．この 8 つ全ての式の両辺を足し合わせると，

$$
\begin{gathered}
\sum_{ijk} n_{ijk} - 4\lambda_1 p_{111}p_{221}p_{212}p_{122} + 4\lambda_1 p_{112}p_{222}p_{211}p_{121} - \lambda_2 \sum_{ijk} p_{ijk} = 0 \\
n_{\cdots} - 4\lambda_1 (p_{111}p_{221}p_{212}p_{122} - p_{112}p_{222}p_{211}p_{121}) - \lambda_2 = 0 \\
\therefore \lambda_2 = n_{\cdots} \quad (\because \text{条件 (2.1)})
\end{gathered}
$$

となる．いま求めた $\lambda_2 = n_{\cdots}$ を式 (2.3) に代入すると，$n_{\cdots} p_{ijk} = m_{ijk}$ より，

$$
\begin{aligned}
&n_{111} - \lambda_1 p_{111}p_{221}p_{212}p_{122} - m_{111} = 0 \qquad (2.4) \\
&n_{221} - \lambda_1 p_{111}p_{221}p_{212}p_{122} - m_{221} = 0 \\
&n_{212} - \lambda_1 p_{111}p_{221}p_{212}p_{122} - m_{212} = 0 \\
&n_{122} - \lambda_1 p_{111}p_{221}p_{212}p_{122} - m_{122} = 0 \\
&n_{112} + \lambda_1 p_{112}p_{222}p_{211}p_{121} - m_{112} = 0 \qquad (2.5) \\
&n_{222} + \lambda_1 p_{112}p_{222}p_{211}p_{121} - m_{222} = 0 \\
&n_{211} + \lambda_1 p_{112}p_{222}p_{211}p_{121} - m_{211} = 0 \\
&n_{121} + \lambda_1 p_{112}p_{222}p_{211}p_{121} - m_{121} = 0
\end{aligned}
$$

となる．ここで，２つの式 (2.4),(2.5) の両辺を足し合わせると，

$$n_{111} - \lambda_1 p_{111}p_{221}p_{212}p_{122} - m_{111} + n_{112} + \lambda_1 p_{112}p_{222}p_{211}p_{121} - m_{112} = 0$$
$$\therefore n_{11\cdot} = m_{11\cdot}$$

となる．同様に適当に式を２つ選び足し合わせると，

$$n_{ij\cdot} = m_{ij\cdot},\ n_{i\cdot k} = m_{ij\cdot},\ n_{\cdot jk} = m_{\cdot jk}$$

が示される．

以上より，m_{ijk} の MLE$\hat{m}_{ijk}$ について，

$$\hat{m}_{ij\cdot} = n_{ij\cdot},\ \hat{m}_{ij\cdot} = n_{i\cdot k},\ \hat{m}_{\cdot jk} = n_{\cdot jk}$$

が成り立つことがわかる． □

また，このモデルのもとでは，

$$\frac{m_{111}m_{ij1}}{m_{i11}m_{1j1}} = \frac{m_{11k}m_{ijk}}{m_{i1k}m_{1jk}} \quad \text{for all } i = 2,\cdots,I,\ j = 2,\cdots,J,\ k = 2,\cdots,K$$

も成り立つ．

本節のはじめに，$M^{(8)}$ の条件式を

$$\frac{p_{111}p_{ij1}}{p_{i11}p_{1j1}} = \frac{p_{11k}p_{ijk}}{p_{i1k}p_{1jk}} \quad \text{(属性 } C \text{ を固定したもとでのオッズ比)}$$

と表記したが，これは

$$\frac{p_{111}p_{i1k}}{p_{i11}p_{11k}} = \frac{p_{1j1}p_{ijk}}{p_{ij1}p_{1jk}} \quad \text{(属性 } B \text{ を固定したもとでのオッズ比)} \tag{2.6}$$

もしくは

$$\frac{p_{111}p_{1jk}}{p_{1j1}p_{11k}} = \frac{p_{i11}p_{ijk}}{p_{ij1}p_{i1k}} \quad \text{(属性 } A \text{ を固定したもとでのオッズ比)} \tag{2.7}$$

と書いても意味は同じである．（この２式は，元の条件式の両辺に $\frac{p_{1j1}p_{i1k}}{p_{ij1}p_{11k}}$ をかけると式 (2.6) が，$\frac{p_{1jk}p_{i11}}{p_{11k}p_{ij1}}$ をかけると式 (2.7) が導かれる．）

2.1.5 ３元分割表モデルのオッズ比表現

本節で挙げた３元分割表モデルは，オッズ比による特徴付けが可能である．

はじめに，条件付独立モデル $M^{(5)}$,$M^{(6)}$,$M^{(7)}$ から記述する．

命題 2.1.2. $M^{(5)}$ が真 $\iff \dfrac{p_{ijk}p_{i'j'k}}{p_{ij'k}p_{i'jk}} = 1 \quad$ for all i, i', j, j', k

命題 2.1.3. $M^{(6)}$ が真 $\iff \dfrac{p_{ijk}p_{i'jk'}}{p_{ijk'}p_{i'jk}} = 1 \quad$ for all i, i', k, k', j

命題 2.1.4. $M^{(7)}$ が真 $\iff \dfrac{p_{ijk}p_{ij'k'}}{p_{ijk'}p_{ij'k}} = 1 \quad$ for all j, j', k, k', i

命題 2.1.2 の証明のみ行う.

(証明)

($\Longrightarrow$) モデル $M^{(5)}$ が真とすると,

$$\begin{aligned}\frac{p_{ijk}p_{i'j'k}}{p_{ij'k}p_{i'jk}} &= \frac{(p_{i\cdot k}p_{\cdot jk}/p_{\cdot\cdot k})(p_{i'\cdot k}p_{\cdot j'k}/p_{\cdot\cdot k})}{(p_{i\cdot k}p_{\cdot j'k}/p_{\cdot\cdot k})(p_{i'\cdot k}p_{\cdot jk}/p_{\cdot\cdot k})} \\ &= \frac{p_{i\cdot k}p_{\cdot jk}p_{i'\cdot k}p_{\cdot j'k}}{p_{i\cdot k}p_{\cdot j'k}p_{i'\cdot k}p_{\cdot jk}} \\ &= 1\end{aligned}$$

となる.

($\Longleftarrow$) 任意の i, i', j, j', k に対して, $(p_{ijk}p_{i'j'k})/(p_{ij'k}p_{i'jk}) = 1$ であるとき, $p_{ijk}p_{i'j'k} = p_{ij'k}p_{i'jk}$ である. これより, 次のようになる.

$$\begin{aligned}p_{ijk}p_{\cdot\cdot k} &= \sum_{i'j'} p_{ijk}p_{i'j'k} \\ &= \sum_{i'j'} p_{ij'k}p_{i'jk} \\ &= \sum_{j'} p_{ij'k} \sum_{i'} p_{i'jk} \\ &= p_{i\cdot k}p_{\cdot jk} \\ \therefore p_{ijk} &= \frac{p_{i\cdot k}p_{\cdot jk}}{p_{\cdot\cdot k}}\end{aligned}$$

□

命題 2.1.3, 2.1.4 もこれと同様に示される.

次に周辺独立モデル $M^{(2)}$,$M^{(3)}$,$M^{(4)}$ について記述する. 先に補題を与える.

補題 2.1.1. $M^{(2)}$ が真 $\iff$ $M^{(5)}$ と $M^{(6)}$ がともに真

補題 2.1.2. $M^{(3)}$ が真 $\iff$ $M^{(5)}$ と $M^{(7)}$ がともに真

補題 2.1.3. $M^{(4)}$ が真 $\iff$ $M^{(6)}$ と $M^{(7)}$ がともに真

補題 2.1.1 の証明のみ行う.

(証明)

($\Longrightarrow$) $M^{(2)}$ が真とすると，$p_{ijk} = p_{i\cdot\cdot}p_{\cdot jk}$ である．このとき，$p_{i\cdot k} = p_{i\cdot\cdot}p_{\cdot\cdot k}$ であるから，$p_{i\cdot\cdot} = p_{i\cdot k}/p_{\cdot\cdot k}$ となる．これより，

$$p_{ijk} = p_{i\cdot\cdot}p_{\cdot jk} = \frac{p_{i\cdot k}p_{\cdot jk}}{p_{\cdot\cdot k}}$$

であり，つまり $M^{(5)}$ を保つ.

同様に，$p_{ijk} = p_{i\cdot\cdot}p_{\cdot jk}$ より $p_{ij\cdot} = p_{i\cdot\cdot}p_{\cdot j\cdot}$ であるから，$p_{i\cdot\cdot} = p_{ij\cdot}/p_{\cdot j\cdot}$ である．これより

$$p_{ijk} = p_{i\cdot\cdot}p_{\cdot jk} = \frac{p_{ij\cdot}p_{\cdot jk}}{p_{\cdot j\cdot}}$$

であり，$M^{(6)}$ も保つことが示される.

($\Longleftarrow$) $M^{(5)}$,$M^{(6)}$ がともに真ならば,

$$p_{ijk} = \frac{p_{i\cdot k}p_{\cdot jk}}{p_{\cdot\cdot k}}, \quad p_{ijk} = \frac{p_{ij\cdot}p_{\cdot jk}}{p_{\cdot j\cdot}}$$

である．この 2 式より,

$$\frac{p_{ijk}}{p_{\cdot jk}} = \frac{p_{i\cdot k}}{p_{\cdot\cdot k}} = \frac{p_{ij\cdot}}{p_{\cdot j\cdot}}$$

である．右側の等式より，$p_{i\cdot k}p_{\cdot j\cdot} = p_{\cdot\cdot k}p_{ij\cdot}$ であるから,

$$\sum_j p_{i\cdot k}p_{\cdot j\cdot} = \sum_j p_{\cdot\cdot k}p_{ij\cdot}$$

$$p_{i\cdot k} = p_{\cdot\cdot k}p_{i\cdot\cdot}$$

$$\therefore \ \frac{p_{i\cdot k}}{p_{\cdot\cdot k}} = p_{i\cdot\cdot}$$

となる．$M^{(5)}$ の式にこれを代入すると,

$$\begin{aligned} p_{ijk} &= \frac{p_{i\cdot k}p_{\cdot jk}}{p_{\cdot\cdot k}} \\ &= p_{i\cdot\cdot}p_{\cdot jk} \end{aligned}$$

となり，$M^{(2)}$ を保つ. □

補題 2.1.2，2.1.3 も同様に示される.

命題 2.1.2，2.1.3，2.1.4 と，補題 2.1.1，2.1.2，2.1.3 から，周辺独立モデル $M^{(2)}$,$M^{(3)}$,$M^{(4)}$ のオッズ比による特徴づけができる.

命題 2.1.5. $M^{(2)}$ が真 $\Longleftrightarrow$

$$\frac{p_{ijk}p_{i'j'k}}{p_{ij'k}p_{i'jk}} = 1 \quad \text{for all} \quad i, i', j, j', k$$

$$\frac{p_{ijk}p_{i'jk'}}{p_{ijk'}p_{i'jk}} = 1 \quad \text{for all} \quad i, i', k, k', j$$

命題 2.1.6. $M^{(3)}$ が真 $\Longleftrightarrow$

$$\frac{p_{ijk}p_{i'j'k}}{p_{ij'k}p_{i'jk}} = 1 \quad \text{for all} \quad i, i', j, j', k$$
$$\frac{p_{ijk}p_{ij'k'}}{p_{ijk'}p_{ij'k}} = 1 \quad \text{for all} \quad j, j', k, k', i$$

命題 2.1.7. $M^{(4)}$ が真 $\Longleftrightarrow$

$$\frac{p_{ijk}p_{i'jk'}}{p_{ijk'}p_{i'jk}} = 1 \quad \text{for all} \quad i, i', k, k', j$$
$$\frac{p_{ijk}p_{ij'k'}}{p_{ijk'}p_{ij'k}} = 1 \quad \text{for all} \quad j, j', k, k', i$$

最後に完全独立モデルについて記述する.

補題 2.1.4. $M^{(1)}$ が真 $\Longleftrightarrow$ $M^{(2)}$,$M^{(3)}$,$M^{(4)}$ 全て真

命題 2.1.8. $M^{(1)}$ が真 $\Longleftrightarrow$

$$\frac{p_{ijk}p_{i'j'k}}{p_{ij'k}p_{i'jk}} = 1 \quad \text{for all} \quad i, i', j, j', k$$
$$\frac{p_{ijk}p_{i'jk'}}{p_{ijk'}p_{i'jk}} = 1 \quad \text{for all} \quad i, i', k, k', j$$
$$\frac{p_{ijk}p_{ij'k'}}{p_{ijk'}p_{ij'k}} = 1 \quad \text{for all} \quad j, j', k, k', i$$

(証明) 補題 2.1.4

($\Longrightarrow$) $M^{(1)}$ が真とすると，$p_{ijk} = p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k}$ である．この式の両辺を i, j, k でそれぞれ和をとると，

$$\begin{cases} p_{\cdot jk} = p_{\cdot j\cdot}p_{\cdot\cdot k} \\ p_{i\cdot k} = p_{i\cdot\cdot}p_{\cdot\cdot k} \\ p_{ij\cdot} = p_{i\cdot\cdot}p_{\cdot j\cdot} \end{cases}$$

である．得られた式をもとの $p_{ijk} = p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k}$ に代入すると，

$$\begin{cases} p_{ijk} = p_{i\cdot\cdot}p_{\cdot jk} \\ p_{ijk} = p_{\cdot j\cdot}p_{i\cdot k} \\ p_{ijk} = p_{\cdot\cdot k}p_{ij\cdot} \end{cases}$$

となり，$M^{(2)}$,$M^{(3)}$,$M^{(4)}$ を保っていることを示している．

($\Longleftarrow$) $M^{(2)}$,$M^{(3)}$,$M^{(4)}$ が全て真であるとする．

$M^{(2)}$: $p_{ijk} = p_{i\cdot\cdot}p_{\cdot jk}$ より，$p_{i\cdot k} = p_{i\cdot\cdot}p_{\cdot\cdot k}$ である．この等式と，$M^{(3)}$: $p_{ijk} = p_{\cdot j\cdot}p_{i\cdot k}$ から，$p_{ijk} = p_{\cdot j\cdot}p_{i\cdot\cdot}p_{\cdot\cdot k}$ を得る．これは $M^{(1)}$ である．

（$M^{(2)}$,$M^{(4)}$ と選んでも，$M^{(3)}$,$M^{(4)}$ と選んでも同様に $M^{(1)}$ が導かれる.） □

補題 2.1.4 より命題 2.1.8 が成立する．

なお，2 元分割表のときと同様，p を m に置き換えることがある．

2.2 ３元分割表の対数線形モデル

２元分割表で用いられた対数線形モデルを３元分割表に拡張する．３元分割表の飽和モデルの形式としては，次が自然である．

$$M_*^{(F)}:\ \log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)}$$

このモデルは，３つの属性でとりうる全ての効果（各主効果, 各２因子交互作用, 各３因子交互作用）を備えたものである．ただし，総セル数 $I \times J \times K$ に対し，右辺のパラメータ数は $1 + I + J + K + IJ + IK + JK + IJK$ 個となり，表のセル数 IJK を超過している．そこで，２元分割表と同様，次のような自然な制約を右辺のパラメータに課す．

$$\begin{cases} \sum_{i=1}^{I} u_{1(i)} = \sum_{j=1}^{J} u_{2(j)} = \sum_{k=1}^{K} u_{3(k)} = 0 \\ \sum_{i=1}^{I} u_{12(ij)} = \sum_{j=1}^{J} u_{12(ij)} = 0 \\ \sum_{i=1}^{I} u_{13(ik)} = \sum_{k=1}^{K} u_{13(ik)} = 0 \\ \sum_{j=1}^{J} u_{23(jk)} = \sum_{k=1}^{K} u_{23(jk)} = 0 \\ \sum_{i=1}^{I} u_{123(ijk)} = \sum_{j=1}^{J} u_{123(ijk)} = \sum_{k=1}^{K} u_{123(ijk)} = 0 \end{cases}$$

この制約により自由なパラメータ数は,

$$\begin{aligned} &1 + (I-1) + (J-1) + (K-1) \\ &+ (I-1)(J-1) + (I-1)(K-1) + (J-1)(K-1) + (I-1)(J-1)(K-1) \\ &= IJK \end{aligned}$$

となり，セル数と一致，つまり一意に定まることになる．

第 2.1 節にて挙げた３元分割表のモデルは，すべて対数線形モデルの形式で書ける．モデル $M^{(x)}$ と同値な対数線形モデルを $M_*^{(x)}$ とし，下に挙げる．

$M_*^{(1)}$: $\log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)}$（完全独立モデル）

$M_*^{(2)}$: $\log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{23(jk)}$（$A \perp\!\!\!\perp (B, C)$ モデル）

$M_*^{(3)}$: $\log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)}$（$B \perp\!\!\!\perp (A, C)$ モデル）

$M_*^{(4)}$: $\log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)}$（$C \perp\!\!\!\perp (A, B)$ モデル）

$M_*^{(5)}$: $\log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)} + u_{23(jk)}$（$A \perp\!\!\!\perp B \mid C$ モデル）

$M_*^{(6)}$: $\log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{23(jk)}$（$A \perp\!\!\!\perp C \mid B$ モデル）

$M_*^{(7)}$: $\log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)}$（$B \perp\!\!\!\perp C \mid A$ モデル）

$M_*^{(8)}$: $\log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)}$（オッズ比均一）

$M^{(x)}$ と $M_*^{(x)}$ が同値であることを示す方法は，2 元分割表のときの方法と同様である．例として，ここでは $M^{(5)}$ と $M_*^{(5)}$ が同値であることを示しておく．

例 2.2.1. $M^{(5)} \Longleftrightarrow M_*^{(5)}$

$M^{(5)}$ が真のとき，

$$\begin{aligned}\log(m_{ijk}) &= \log(n_{\cdots}p_{ijk}) \\ &= \log\left(n_{\cdots}\cdot\frac{p_{i\cdot k}p_{\cdot jk}}{p_{\cdot\cdot k}}\right) \\ &= \log(n_{\cdots}) - \log(p_{\cdot\cdot k}) + \log(p_{i\cdot k}) + \log(p_{\cdot jk})\end{aligned}$$

となるので，$u_{1(i)} = u_{2(j)} = 0$ ととれば，これは $M_*^{(5)}$ の形式であることがわかる．

逆に，$M_*^{(5)}$ が真のとき

$$\begin{aligned}&\log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)} + u_{23(jk)} \\ \Longleftrightarrow\quad & m_{ijk} = \exp\left\{u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)} + u_{23(jk)}\right\} \\ &= aa_{1(i)}a_{2(j)}a_{3(k)}a_{13(ik)}a_{23(jk)} \\ &\quad (e^u = a, e^{u_{1(i)}} = a_{1(i)}, \ldots, e^{u_{23(jk)}} = a_{23(jk)})\end{aligned}$$

である．このとき，$p_{ijk} = m_{ijk}/n_{\cdots}$ より，

$$\begin{aligned}&p_{ijk} = \frac{aa_{1(i)}a_{2(j)}a_{3(k)}a_{13(ik)}a_{23(jk)}}{n_{\cdots}} \\ \Longleftrightarrow\ &\begin{cases} p_{i\cdot k} = \dfrac{aa_{1(i)}a_{2(\cdot)}a_{3(k)}a_{13(ik)}a_{23(\cdot k)}}{n_{\cdots}} \\ p_{\cdot jk} = \dfrac{aa_{1(\cdot)}a_{2(j)}a_{3(k)}a_{13(\cdot k)}a_{23(jk)}}{n_{\cdots}} \end{cases}\end{aligned}$$

である．これより，

$$\begin{aligned}p_{i\cdot k}p_{\cdot jk} &= \frac{aa_{1(i)}a_{2(\cdot)}a_{3(k)}a_{13(ik)}a_{23(\cdot k)}}{n_{\cdots}} \cdot \frac{aa_{1(\cdot)}a_{2(j)}a_{3(k)}a_{13(\cdot k)}a_{23(jk)}}{n_{\cdots}} \\ &= \frac{aa_{1(i)}a_{2(j)}a_{3(k)}a_{13(ik)}a_{23(jk)}}{n_{\cdots}} \cdot \frac{aa_{1(\cdot)}a_{2(\cdot)}a_{3(k)}a_{13(\cdot k)}a_{23(\cdot k)}}{n_{\cdots}} \\ &= p_{ijk}p_{\cdot\cdot k}\end{aligned}$$

$$\therefore p_{ijk} = \frac{p_{i\cdot k}p_{\cdot jk}}{p_{\cdot\cdot k}} \text{となり} M^{(5)} \text{が真となる．}$$

対数線形モデルにおいて，ある交互作用が存在するときは，その交互作用を構成する属性同士は独立でない．

例 2.2.2. モデル $M_*^{(4)}$ は $\log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)}$ である．この式から，

$$\begin{aligned}&m_{ijk} = \exp\{u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)}\} \\ \Longleftrightarrow\ &n_{\cdots}p_{ijk} = \exp\{u_{3(k)}\}\exp\{u + u_{1(i)} + u_{2(j)} + u_{12(ij)}\}\end{aligned}$$

$$\Longleftrightarrow p_{ijk} = \frac{1}{n_{...}} \exp\{u_{3(k)}\} \exp\{u + u_{1(i)} + u_{2(j)} + u_{12(ij)}\}$$

のように，$p_{ijk} = f(i,j)g(k)$ の形の式が導かれる．
この式から，(i,j) と k，つまり属性 (A,B) と C は独立であることがわかる．また，i と j，つまり属性 A と B は独立でないことがわかる．

また，$u_{12(ij)}$ と $u_{13(ik)}$ のように，２つの交互作用の構成にある１つの属性が共通しているときは，共通の属性を与えたもとで残り２つの属性は条件付独立である．

例 2.2.3. モデル $M_*^{(7)}$ は $\log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)}$ ある．この式から，

$$\begin{aligned} m_{ijk} &= \exp\{u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)}\} \\ \Longleftrightarrow n_{...}p_{ijk} &= \exp\{u + u_{1(i)} + u_{2(j)} + u_{12(ij)}\} \exp\{u_{3(k)} + u_{13(ik)}\} \\ \Longleftrightarrow p_{ijk} &= \frac{1}{n_{...}} \exp\{u + u_{1(i)} + u_{2(j)} + u_{12(ij)}\} \exp\{u_{3(k)} + u_{13(ik)}\} \end{aligned}$$

のように，$p_{ijk} = f(i,j)g(i,k)$ の形の式が導かれる．**因数分解基準**より，i を与えたもとで j,k は独立，つまり $B \perp\!\!\!\perp C|A$ となる．

以上のように，モデルが含む交互作用は，属性間の関係を示している．

定義 2.2.1. 条件「ある交互作用がモデルが含まれるなら，それに含まれるより低次の交互作用および主効果項をモデルに入れる」を満たすモデルを**階層モデル**とよぶ．

階層モデルは自然なモデルであり，解析者にとっても望ましいものであるといえる．例えば，あるモデルに $u_{12(ij)}$ が存在するのに，それより小さい効果である $u_{1(i)}, u_{2(j)}$ は存在しないと仮定するのは不自然で，解釈も困難である．

以降，考察する対象をこの階層モデルに限定する．

階層モデルでは，モデルに含まれる極大な交互作用もしくは主効果項による簡潔なモデル表現が可能である．例えば，本節の始めに挙げた３元分割表の飽和モデルには，３因子交互作用が存在し，階層モデルの定義から低次の効果はすべて含まれる．よって，３因子交互作用 $u_{123(ijk)}$ が極大項である．そこでこのモデルを，属性 1,2,3 の３因子交互作用が極大であるという意味で，[123] と表現することにする．$M_*^{(5)}$ の場合，極大な項は $u_{13(ik)}, u_{23(jk)}$ である．これを [13][23] と表現することにする．以上のように，極大な項でそのモデルを表現する．

本節で挙げたモデルについて，極大項と簡潔な表現（速記）をまとめた．

	極大項	速記
$M_*^{(1)}$	$u_{1(i)}, u_{2(j)}, u_{3(k)}$	[1][2][3]
$M_*^{(2)}$	$u_{1(i)}, u_{23(jk)}$	[1][23]
$M_*^{(3)}$	$u_{2(j)}, u_{13(ik)}$	[2][13]
$M_*^{(4)}$	$u_{3(k)}, u_{12(ij)}$	[3][12]
$M_*^{(5)}$	$u_{13(ik)}, u_{23(jk)}$	[13][23]
$M_*^{(6)}$	$u_{12(ij)}, u_{23(jk)}$	[12][23]
$M_*^{(7)}$	$u_{12(ij)}, u_{13(ik)}$	[12][13]
$M_*^{(8)}$	$u_{12(ij)}, u_{13(ik)}, u_{23(jk)}$	[12][13][23]
飽和モデル	$u_{123(ijk)}$	[123]

極大項の集合のことを**生成集合**とよぶ．またこのモデルの速記により，属性間の関係も判別しやすくなる．例えば，[1][23] は 1 と 23 が異なる括弧で区切られているから，この 2 つの組が独立であるということを説明していることになるし，[12][13] は，1 が共通で 2,3 が別々の括弧で区切られているから，1 を与えたもとで 2,3 は独立であることを説明している．これは，前の交互作用に関する記述からもわかる．

2.3 モデル検定

3 元分割表モデル M_* の適合度検定を考える．ここでの仮説というのは,

H_0 : モデル M_*は真
H_1 : モデル M_*は偽　（当てはまるのは飽和モデル）

である．2 元分割表分析で利用した 2 つの検定統計量を，3 元分割表に拡張する．

- カイ二乗検定統計量：

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\frac{(n_{ijk}-m_{ijk})^2}{m_{ijk}}$$

- 尤度比検定統計量：

$$G^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} n_{ijk}\frac{n_{ijk}}{m_{ijk}}$$

ここで，m_{ijk} は M_* のもとでの期待度数の MLE である．この 2 つは，飽和モデルに対する仮説モデルの検定の検定統計量である．自由度は，対立仮説と帰無仮説，各場合での自由なパラメータの数の差である．

例 2.3.1. 飽和モデルに対するモデル $M_*^{(1)}$ の検定を考える．飽和モデルに存在するが $M_*^{(1)}$ に存在しないパラメータは,

$$u_{12(ij)}, u_{13(ik)}, u_{23(jk)}, u_{123(ijk)}$$

の 4 種類である．制約を考慮すると，そのうちの自由なパラメータの総数は，

$$\begin{aligned}
&(I-1)(J-1)+(I-1)(K-1)+(J-1)(K-1)+(I-1)(J-1)(K-1)\\
&=\{IJ-I-J+1\}+\{IK-I-K+1\}+\{JK-J-K+1\}\\
&\quad+\{IJK-IJ-IK-JK+I+J+K-1\}\\
&=IJK-I-J-K+2
\end{aligned}$$

である．これより，尤度比検定統計量

$$G^2=2\sum_{ijk}n_{ijk}\log\left(\frac{n_{ijk}}{\hat{m}^{(1)}_{ijk}}\right)$$

の自由度は $IJK-I-J-K+2$ である．

分割表解析においては，G^2 が扱いやすい．その理由は，飽和モデルに対する検定統計量 G^2 を用いて，大小関係にあるモデル同士の比較の検定ができるからである．ここで,「M_* に比べ大きいモデル」とは，M_* のもつ項をすべて含むモデルのことである．例えば，[1][23] と [12][23], [1][2][3] と [12][23] のような関係である．[12][13][23] は飽和モデルを除くすべてに対して相対的に大きい．このような大小関係が成り立つモデル同士は比較可能である．

比較可能ではないのは，例えば，[2][13] と [12][23] のような大小関係がない場合である．

モデル $M^{(r)}_*$ とモデル $M^{(s)}_*$ があるとし，$M^{(s)}_*$ が相対的に大きいモデルとする．これらのモデルの飽和モデルに対する尤度比検定統計量は，

$$G^2_r=2\sum_{ijk}n_{ijk}\log\left(\frac{n_{ijk}}{\hat{m}^{(r)}_{ijk}}\right)$$

$$G^2_s=2\sum_{ijk}n_{ijk}\log\left(\frac{n_{ijk}}{\hat{m}^{(s)}_{ijk}}\right)$$

である． 2 式の差をとると，

$$G^2_r-G^2_s=2\sum_{ijk}n_{ijk}\log\left(\frac{\hat{m}^{(s)}_{ijk}}{\hat{m}^{(r)}_{ijk}}\right)$$

であり，第 3 章で述べるように，対数線形モデルの場合，この右辺は

$$G^2_r-G^2_s=2\sum_{ijk}\hat{m}^{(s)}_{ijk}\log\left(\frac{\hat{m}^{(s)}_{ijk}}{\hat{m}^{(r)}_{ijk}}\right)$$

と変形できる．これは，第 2 章で紹介した形であり，$M^{(r)}_*$ の $M^{(s)}_*$ に対する尤度比検定統計量である．

$$G^2(M^{(r)}_*vs\ M^{(s)}_*)=G^2_r-G^2_s,\quad M^{(r)}_*<M^{(s)}_*$$

自由度も同じく２つの G^2 の自由度の差であり，

$$df\left(G^2(M_*^{(r)} vs\ M_*^{(s)})\right) = df\left(G_r^2\right) - df\left(G_s^2\right)$$

となる．これは，

$$df\left(G_r^2\right) = df(M_*^{(F)}) - df(M_*^{(r)})$$
$$df\left(G_s^2\right) = df(M_*^{(F)}) - df(M_*^{(s)})$$

より，

$$df\left(G_r^2\right) - df\left(G_s^2\right) = df(M_*^{(s)}) - df(M_*^{(r)})$$

となり，確かに２つのモデルの自由度の差となっている．この「差の統計量」は，漸近的に

$$G^2(M_*^{(r)} vs M_*^{(s)}) \sim \chi^2(df(M_*^{(s)}) - df(M_*^{(r)}))$$

が成り立つため，新たな検定統計量として用いることができる．

このように，対飽和モデルの検定の結果から縮小モデルの検討が行えるという点で，G^2 は有用で，良く利用される．

また，この検定法は，効果項の有無に関する検定と言い換えることができる．例えば [1][23] に対する [12][23] の比較検定は，

$$\mathrm{H}_0 : \quad u_{12(ij)} = 0 \text{ for all } i, j$$
$$\mathrm{H}_1 : \quad u_{12(ij)} \neq 0 \text{ for some } i, j$$

という検定と同値である．

2.4　積多項サンプリング

積多項サンプリング表の場合について記述する．積多項表では，総度数とは別に周辺度数が固定されていた．積多項表モデルの検討は，この周辺度数の制約を満たすモデルに制限されることになる．

例 2.4.1. 次の表は，10 代・20 代・30 代の女性と男性に対し，ある質問を行い，その解答結果を性別，年代で分類してまとめたものである．

		解答 (k)		
性別 (i)	年代 (j)	賛成	反対	計
女性	10 代	79	40	119
	20 代	132	71	203
	30 代	98	80	178
	計	309	191	500
男性	10 代	65	94	159
	20 代	141	95	236
	30 代	113	92	205
	計	319	281	600

今回の調査の性質上，この 3 元分割表は $n_{ij.}$ が固定されている積多項表である．つまり，

$$n_{ij.} = m_{ij.}$$

という周辺制約がある．対数線形モデル M_* と同値なモデル M には元々それぞれ周辺制約があった．よって，積多項表にモデルを考えるときは，この制約を満たすものに限定する必要がある．今回は $n_{ij.} = m_{ij.}$ の制約があるため，それを満たすモデル

$$\begin{cases} M_*^{(4)}, [12][3] \\ M_*^{(6)}, [12][23] \\ M_*^{(7)}, [12][13] \\ M_*^{(8)}, [12][13][23] \\ M_*^{(F)}, [123] \end{cases}$$

に制限する．
また，別の考え方で，モデルの制限を行うこともできる．今の表は $n_{ij.}$ が固定されているから，各セルの期待度数は $m_{ijk} = n_{ij.}p_{ijk}$ である．この式より，$\log(m_{ijk}) = \log(n_{ij.}) + \log(p_{ijk})$ であるから，この表の対数線形モデルには，i, j のみに依存する効果項 $u_{12(ij)}$ が含まれるべきである．この条件からも，今挙げたモデルに制限できる．

本節での積多項表の扱いは，高次元の表でも同様である．

2.5 分割表併合

対象のデータの変数の数が多い場合，解析者は可能なら変数を減らしてしまいたい．変数を減らすということは，分割表を併合（縮約）することを意味する．

シンプソンのパラドックス

例 2.5.1. 次の表は，ある 390 人の患者を，性別，治療法（A,B），治療結果（成功, 失敗）の 3 つの属性で分類した 3 元分割表データである．

		性別			
		男性		女性	
治療結果		成功	失敗	成功	失敗
治療法	A	60	20	40	80
	B	100	50	10	30

男性側の治療法と結果の表に注目する．治療 A での成功の確率は，$60/80 = 0.75$ と推定される．治療 B での成功の確率は，$100/150 = 0.667$ と推定される．これより，男性の場合であると，治療 1 の方が成功の可能性が高いと考えられる．

同様に女性側の表についても考える．治療 A での成功の確率は，$40/120 = 0.333$ と推定される．治療 B での成功の確率は，$10/40 = 0.25$ と推定される．これより，男性と同様，女性でも治療 2 の方が成功の可能性が高いと考えられる．

ここで，男性の表と女性の表を併合することで得る治療法と結果の周辺度数表について考えてみる．

		結果	
		成功	失敗
治療法	A	100	100
	B	110	80

この周辺表についても同様に考えると，治療 A での成功の確率 p_{11} は，$100/200 = 0.5$ と推定される．一方，治療 B での成功の確率 p_{21} は，$110/190 = 0.579$ と推定される．
これより治療 B の方が可能性が高いという先とは異なる結果となってしまった．このような表の併合で起こる矛盾のことを，**シンプソンのパラドックス**という．

周辺度数の差により生じるシンプソンパラドックスの可能性があるため，安易に多元分割表を縮約して元より小さい次元の表にすることはできない．ただし，条件を満たせば，併合可能である．

例 2.5.2. $I \times J \times K$ の 3 元分割表で考える．. 属性 C の第 k 水準のもとでの $I \times J$ 分割表の全セル度数の割合ベクトルを

$$\boldsymbol{p}_k = (p_{11k}, \ldots, p_{IJk})' \quad k = 1, \ldots, K, \ \left(p_{ijk} = \frac{n_{ijk}}{n_{\cdot\cdot k}} \right)$$

とすると，この割合ベクトルが任意の C_k について等しい，つまり

$$\boldsymbol{p}_k = (p_{11k}, \ldots, p_{IJk})' = \left(p_{11k}^{(0)}, \ldots, p_{IJk}^{(0)} \right)' \quad k = 1, \ldots, K$$

ときに，表の併合が可能である．属性 A と B の頻度の分割表が属性 C の水準によらず等しい場合，属性 C について併合が可能であり，$I \times J$ 表に縮約できる．

2.6 データ解析例

次のデータは，平成 23 年に行われた外交に関する世論調査における「アメリカに対して親しみを感じますか」という質問への回答を性別，年齢で分類したものである．なお，こちらのデータは，資料 [9]（内閣府ホームページ）から引用したものを解析用に修正している．

Q．アメリカに対して親しみを感じますか ？

	男性			女性		
	はい	いいえ	小計	はい	いいえ	小計
２０～２９歳	69	14	83	66	9	75
３０～３９歳	98	22	120	130	29	159
４０～４９歳	117	15	132	125	21	146
５０～５９歳	129	25	154	131	25	156
６０～６９歳	169	24	193	198	34	232
７０歳以上	183	30	213	152	48	200

３つの属性に属性番号 1 : 性別, 2 : 年齢, 3 : 回答 と振り分けると（対応する添字は，順に i, j, k），この表は $2 \times 6 \times 2$ 分割表であり，データの性質から，$n_{ij\cdot}$ が固定の積多項サンプリング表とみなせる．各添字は，ここでは

$$i = \begin{cases} 1 : \text{男性} \\ 2 : \text{女性} \end{cases} \quad j = \begin{cases} 1 : 20\text{〜}29\text{ 歳} \\ \vdots \\ 6 : 70\text{ 以上歳} \end{cases} \quad k = \begin{cases} 1 : \text{はい} \\ 2 : \text{いいえ} \end{cases}$$

と対応させる．

この積多項表に，条件 $\hat{m}_{ij\cdot} = n_{ij\cdot}$ を満たすモデル $M_*^{(4)}, M_*^{(6)}, M_*^{(7)}, M_*^{(8)}, M_*^{(F)}$ を R を用いて当てはめる．

なお，R の関数 loglm でも，モデルの指定は極大項で可能である．

```
> rdata <- read.csv("america.csv",header=T)
> rdata
   Sex Age Answer   N
1    M  A2      Y  69
2    M  A2      N  14
3    M  A3      Y  98
4    M  A3      N  22
5    M  A4      Y 117
6    M  A4      N  15
7    M  A5      Y 129
8    M  A5      N  25
9    M  A6      Y 169
10   M  A6      N  24
11   M  A7      Y 183
12   M  A7      N  30
13   F  A2      Y  66
14   F  A2      N   9
```

```
15   F  A3      Y 130
16   F  A3      N  29
17   F  A4      Y 125
18   F  A4      N  21
19   F  A5      Y 131
20   F  A5      N  25
21   F  A6      Y 198
22   F  A6      N  34
23   F  A7      Y 152
24   F  A7      N  48
>
>#モデル当てはめ（極大項で指定）
> mfull <- loglm(N~Sex*Age*Answer,data=rdata)              #飽和モデル
> m8 <- loglm(N~Sex*Age+Sex*Answer+Age*Answer,data=rdata) #モデル 8
> m7 <- loglm(N~Sex*Age+Sex*Answer,data=rdata)             #モデル 7
> m6 <- loglm(N~Sex*Age+Age*Answer,data=rdata)             #モデル 6
> m4 <- loglm(N~Sex*Age+Answer,data=rdata)                 #モデル 4
>
>#各モデルの当てはめ結果出力
>
> mfull   #飽和モデル:[123]
Call:
loglm(formula = N ~ Sex * Age * Answer, data = rdata)

Statistics:
                 X^2 df P(> X^2)
Likelihood Ratio   0  0        1   #完全に当てはまる
Pearson            0  0        1
>
> m8   #モデル 8:[12][13][23]
Call:
loglm(formula = N ~ Sex * Age + Sex * Answer + Age * Answer,
    data = rdata)

Statistics:
```

```
                     X^2 df  P(> X^2)
Likelihood Ratio 5.872805  5 0.3187930
Pearson          5.841410  5 0.3219576
>
> m7   #モデル 7:[12][13]
Call:
loglm(formula = N ~ Sex * Age + Sex * Answer, data = rdata)

Statistics:
                     X^2 df  P(> X^2)
Likelihood Ratio 13.60425 10 0.1918199
Pearson          13.98503 10 0.1736754
>
> m6   #モデル 6:[12][23]
Call:
loglm(formula = N ~ Sex * Age + Age * Answer, data = rdata)

Statistics:
                    X^2 df  P(> X^2)
Likelihood Ratio 8.42175  6 0.2088037
Pearson          8.37466  6 0.2119193
>
> m4   #モデル 4:[12][3]
Call:
loglm(formula = N ~ Sex * Age + Answer, data = rdata)

Statistics:
                     X^2 df  P(> X^2)
Likelihood Ratio 16.00604 11 0.1409067
Pearson          16.73963 11 0.1158166
```

飽和モデルを除くと，$M_*^{(8)}$ の P 値が他のモデルに比べて大きくなるのは自然のことであるが，一段階小さい（2 因子交互作用が 1 つ少ない）モデルである $M_*^{(6)}$ と $M_*^{(7)}$ も P 値を見る限りそれなりに当てはまりがよいモデルと考えられる．だが，さらに一段階小さい $M_*^{(4)}$ も悪いモデルとはいえないことから，このデータには条件に合う最小のモデル $M_*^{(4)}$ を採用したい．

$$M_*^{(4)} : \log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} \quad ([3][12])$$

以下，このモデル $M_*^{(4)}$ の各パラメータの推定値と，モデルのもとでの当てはめ値を算出する．

```
> m4.glm <- glm(N~Sex*Age+Answer,family=poisson,
                contrasts=list(Sex="contr.sum",Age="contr.sum",
                               Answer="contr.sum"),data=rdata)
> summary(m4.glm)

Call:
glm(formula = N ~ Sex * Age + Answer, family = poisson, data = rdata,
    contrasts = list(Sex = "contr.sum", Age = "contr.sum", Answer = "contr.sum"))

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.37485  -0.46712   0.01213   0.30563   2.67400

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.98661    0.03268 121.977  < 2e-16 ***
Sex1         0.03457    0.02451   1.410   0.1584
Age1        -0.62475    0.06951  -8.988  < 2e-16 ***
Age2        -0.06472    0.05512  -1.174   0.2403
Age3        -0.05971    0.05482  -1.089   0.2760
Age4         0.05049    0.05245   0.963   0.3358
Age5         0.36180    0.04672   7.744 9.64e-15 ***
Answer1     -0.83328    0.03169 -26.296  < 2e-16 ***
Sex1:Age1   -0.08525    0.06951  -1.226   0.2200
Sex1:Age2    0.10614    0.05512   1.926   0.0541 .
Sex1:Age3    0.01583    0.05482   0.289   0.7727
Sex1:Age4   -0.02812    0.05245  -0.536   0.5919
Sex1:Age5    0.05745    0.04672   1.230   0.2188
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1144.763  on 23  degrees of freedom
```

```
Residual deviance:   16.006  on 11  degrees of freedom
AIC: 181.54

Number of Fisher Scoring iterations: 4

> fitted(m4.glm)  #当てはめ値
        1         2         3         4         5         6
 69.81267  13.18733 100.93398  19.06602 111.02738  20.97262
        7         8         9        10        11        12
129.53194  24.46806 162.33548  30.66452 179.15781  33.84219
       13        14        15        16        17        18
 63.08374  11.91626 133.73752  25.26248 122.80301  23.19699
       19        20        21        22        23        24
131.21417  24.78583 195.13902  36.86098 168.22330  31.77670
```

途中で summary 関数で呼び出されたパラメータ推定値だが，今回 R の仕様で，元のデータへの割り振りから変化し，パラメータの表記がややこしくなってしまっている（出力における Sex1 は Sex が F（女性）の主効果であり，Answer1 は Answer が N（いいえ）の主効果である）．ここでは，元のデータ（rdata）の割り振りに対応するよう，わかりやすくパラメータ推定値をまとめた．

効果項	summary での表記	パラメータ	推定値
切片	Intercept	u	3.98661
第 1 主効果	-	$u_{1(1)}$	-0.03457
	Sex1	$u_{1(2)}$	0.03457
第 2 主効果	Age1	$u_{2(1)}$	-0.62475
	Age2	$u_{2(2)}$	-0.06472
	Age3	$u_{2(3)}$	-0.05971
	Age4	$u_{2(4)}$	0.05049
	Age5	$u_{2(5)}$	0.3618
	-	$u_{2(6)}$	0.33689
第 3 主効果	-	$u_{3(1)}$	0.83328
	Answer1	$u_{3(2)}$	-0.83328
$[1-2]$ 交互作用	-	$u_{12(11)}$	0.08525
	-	$u_{12(12)}$	-0.10614
	-	$u_{12(13)}$	-0.01583
	-	$u_{12(14)}$	0.02812
	-	$u_{12(15)}$	-0.05745
	-	$u_{12(16)}$	0.06605
	Sex1:Age1	$u_{12(21)}$	-0.08525
	Sex1:Age2	$u_{12(22)}$	0.10614
	Sex1:Age3	$u_{12(23)}$	0.01583
	Sex1:Age4	$u_{12(24)}$	-0.02812
	Sex1:Age5	$u_{12(25)}$	0.05745
	-	$u_{12(26)}$	-0.06605

以上から，例えば，70 歳以上の女性で，質問に「いいえ」と回答する人数は，$M_*^{(4)}$ のもとでは，

$$\begin{aligned}\log(m_{262}) &= u + u_{1(2)} + u_{2(6)} + u_{3(2)} + u_{12(26)} \\ &= 3.98661 + 0.03457 + 0.33689 - 0.83328 - 0.06605 = 3.45874 \\ &\therefore m_{262} = e^{3.45874} \fallingdotseq 31.77\end{aligned}$$

と推定される．

採択されたモデルは，[3][12] だった．これは，回答は「性別と年齢」に影響を受けないということだ．言い換えると，回答の「はい」と「いいえ」の比率は，全 12 の母集団（性別と年齢の組合せ）間には差がないということである．このことを，性別と年齢の組合せを行に，回答を列に設定した 12×2 の 2 元分割表の解析でも確かめてみる．

	はい	いいえ	小計
男性２０～２９歳	69	14	83
男性３０～３９歳	98	22	120
男性４０～４９歳	117	15	132
男性５０～５９歳	129	25	154
男性６０～６９歳	169	24	193
男性７０歳以上	183	30	213
女性２０～２９歳	66	9	75
女性３０～３９歳	130	29	159
女性４０～４９歳	125	21	146
女性５０～５９歳	131	25	156
女性６０～６９歳	198	34	232
女性７０歳以上	152	48	200

各行の 12 の母集団に差がないとは，第１章で見たとおり，対数線形モデル $\log(m_{ij}) = u + u_{1(i)} + u_{2(j)}$ が当てはまるということである．実際に解析する．

```
> rdata2 <- read.csv("america2.csv",header=T)
> rdata2
    S.A Answer   N
1  M-A2      Y  69
2  M-A2      N  14
3  M-A3      Y  98
4  M-A3      N  22
5  M-A4      Y 117
6  M-A4      N  15
7  M-A5      Y 129
8  M-A5      N  25
9  M-A6      Y 169
10 M-A6      N  24
11 M-A7      Y 183
12 M-A7      N  30
13 F-A2      Y  66
14 F-A2      N   9
15 F-A3      Y 130
16 F-A3      N  29
17 F-A4      Y 125
18 F-A4      N  21
19 F-A5      Y 131
```

```
20 F-A5      N  25
21 F-A6      Y 198
22 F-A6      N  34
23 F-A7      Y 152
24 F-A7      N  48
> model <- loglm(N~S.A+Answer,data=rdata2)
> model
Call:
loglm(formula = N ~ S.A + Answer, data = rdata2)

Statistics:
                      X^2 df  P(> X^2)
Likelihood Ratio 16.00604 11 0.1409067
Pearson          16.73963 11 0.1158166
```

P 値からも $\log(m_{ij}) = u + u_{1(i)} + u_{2(j)}$ が当てはまりが悪くないことがわかる．なお，P 値がモデル $M_*^{(4)}$（[3][12]）と等しいのは，２つのモデルは本質的に同じものであるからである．

第3章　多元分割表

2元分割表, 3元分割表と順に記述してきた．この章では，多元分割表解析について考察する．一般化のために，次のような表記を与える．

- q : 分割表の全セル数
- $\boldsymbol{n} = (n_1, \ldots, n_q)'$: 度数ベクトル
- $\boldsymbol{p} = (p_1, \ldots, p_q)'$: 確率ベクトル
- $\boldsymbol{m} = (m_1, \ldots, m_q)'$: 期待度数ベクトル

3.1　多元分割表の対数線形モデル

ここではモデル化する対象を全 q セルを保有する多元分割表として，対数線形モデルを一般化したい．このために，ベクトルと行列を用いて対数線形モデルを表現する．いくつかの表記を用意しておく．

- r : モデルの母数の数
- $\boldsymbol{\beta}$: 母数ベクトル（r 次列ベクトル)
- X : モデル計画行列（$q \times r$ 行列)

ここでいうモデルの母数とは，モデルを構成する主効果や交互作用のことであり，これを並べたベクトル $\boldsymbol{\beta}$ のことを効果ベクトルともよぶ.
また形式的に，任意のベクトル $\boldsymbol{v}$= $(v_1, \ldots, v_q)'$ に対し，

- $\log(\boldsymbol{v}) = (\log v_1, \ldots, \log v_q)'$

と定めておく.

これらを用意したことで，一般的な対数線形モデルは

$$\log(\boldsymbol{m}) = X\boldsymbol{\beta}$$

と表現できる．また，

$$\boldsymbol{\mu} = \log(\boldsymbol{m})$$

と表すことにする.

例 3.1.1. 2×2 分割表の対数線形モデル $\log(m_{ij}) = u + u_{1(i)} + u_{2(j)} + u_{12(ij)}$ は,

$$\begin{bmatrix} \log(m_{11}) \\ \log(m_{12}) \\ \log(m_{21}) \\ \log(m_{22}) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ u_{1(1)} \\ u_{1(2)} \\ u_{2(1)} \\ u_{2(2)} \\ u_{12(11)} \\ u_{12(12)} \\ u_{12(21)} \\ u_{12(22)} \end{bmatrix}$$

のように書ける.

ここから，多元の多項サンプリング表と積多項サンプリング表についてそれぞれ考察する.

多項サンプリング

推定

多項サンプリング表に対し，対数線形モデル $\log(\boldsymbol{m}) = X\boldsymbol{\beta}$ を仮定する．また，$\sum_{i=1}^{q} p_i = 1$ である.

期待度数の MLE $\hat{\boldsymbol{m}}$ について考える．多項モデルのもとでの尤度関数は,

$$L(\boldsymbol{p}) = \frac{n_{\cdot}!}{\prod_{i=1}^{q} n_i!} \prod_{i=1}^{q} p_i^{n_i}$$

である．この時点では尤度関数は $\boldsymbol{p}$ の関数として書かれているが，多項モデルの $m_i = n_{\cdot} p_i$ という性質により，$\boldsymbol{m}$ の同値な関数として書き換えられる．その関数は,

$$L(\boldsymbol{m}) = \frac{n_{\cdot}!}{\prod_{i=1}^{q} n_i!} \prod_{i=1}^{q} \left(\frac{m_i}{n_{\cdot}} \right)^{n_i}$$

と与えられる．対数変換を行い，$\boldsymbol{m}$ に依存する項のみを残すと,

$$\ell(\boldsymbol{m}) = \sum_{i=1}^{q} n_i \log(m_i) = \boldsymbol{n}' \log(\boldsymbol{m})$$

となる．これを最大にする $\boldsymbol{m}$=$\hat{\boldsymbol{m}}$ が MLE である.

$\hat{\boldsymbol{m}}$ には次の 2 つの制約がある.

- 対数線形構造

$$\log(\hat{\boldsymbol{m}}) = X\hat{\boldsymbol{\beta}} \quad \text{for some} \quad \hat{\boldsymbol{\beta}} \tag{3.1}$$

- 度数総和と期待度数総和が一致

  $\boldsymbol{J}=(1,\ldots,1)'$ (q 次ベクトル) とすると,

$$\boldsymbol{n}'\boldsymbol{J}=\hat{\boldsymbol{m}}'\boldsymbol{J} \quad (n.=m.) \tag{3.2}$$

$\hat{\boldsymbol{m}}$ は以上の条件を満たさなければならない.

また, $\hat{\boldsymbol{m}}$ には独自の性質がある. それに関する次の命題がある.

命題 3.1.1. 多項サンプリングモデルの尤度関数 $L(\boldsymbol{m})$ を最大にする点 $\hat{\boldsymbol{m}}$ は,

$$\hat{\boldsymbol{m}}'X=\boldsymbol{n}'X$$

を満たし, また一意である.

(証明)

前の尤度関数を変形する. 条件 (3.2) より, $\exp(n.)-\exp(m.)=0$ である. よって $L(\boldsymbol{m})$ は,

$$L(\boldsymbol{m})=\frac{n.!}{\prod_{i=1}^{q}n_i!}e^{n.}\prod_{i=1}^{q}\left\{\left(\frac{m_i}{n.}\right)^{n_i}e^{-m_i}\right\}$$

と変形可能である. 対数変換すると,

$$\begin{aligned}\log L(\boldsymbol{m})&=\log\left\{\frac{n.!}{\prod_{i=1}^{q}n_i!}e^{n.}\right\}\log\left\{\prod_{i=1}^{q}\left(\frac{m_i}{n.}\right)^{n_i}e^{-m_i}\right\}\\&=\log\left\{\frac{n.!}{\prod_{i=1}^{q}n_i!}e^{n.}\right\}+\sum_{i=1}^{q}(n_i\log m_i-n_i\log n.-m_i)\end{aligned}$$

となる. $L(\boldsymbol{m})$ を最大にするには, $\boldsymbol{m}$ に依存しない部分を除いた

$$\ell(\boldsymbol{m})=\sum_{i=1}^{q}n_i\log m_i-\sum_{i=1}^{q}m_i$$

を最大化すれば十分である.

ここで, $\boldsymbol{\mu}=\log(\boldsymbol{m})=X\boldsymbol{\beta}$ であったから, $\ell(\boldsymbol{m})$ は $\boldsymbol{\mu}$ の関数であると同時に, $\boldsymbol{\mu}$ も $\boldsymbol{\beta}$ の関数である. $f(\boldsymbol{\beta})$ を

$$f(\boldsymbol{\beta})=\ell(\boldsymbol{\mu})=\sum_{i=1}^{q}n_i\mu_i-\sum_{i=1}^{q}\exp\mu_i$$

とおき, この $f(\boldsymbol{\beta})$ を $\boldsymbol{\beta}$ について微分すると,

$$\begin{aligned}\frac{df}{d\boldsymbol{\beta}}&=\frac{d\ell}{d\boldsymbol{\mu}}\cdot\frac{d\boldsymbol{\mu}}{d\boldsymbol{\beta}}\\&=\frac{d\ell}{d\boldsymbol{\mu}}\left\{\sum_{i=1}^{q}n_i\mu_i-\sum_{i=1}^{q}\exp\mu_i\right\}\cdot\frac{d}{d\boldsymbol{\beta}}(X\boldsymbol{\beta})\\&=\left\{\boldsymbol{n}'-(\exp\mu_1,\ldots,\exp\mu_q)'\right\}X\end{aligned}$$

$$= (\boldsymbol{n}' - \boldsymbol{m}')X$$

となる．$\dfrac{df}{d\boldsymbol{\beta}} = \mathbf{0}$ とおいたときの解 $\boldsymbol{m}$=$\hat{\boldsymbol{m}}$ は，$\ell(\boldsymbol{m})$ を最大，同時に $L(\boldsymbol{m})$ を最大にする．実際に $\dfrac{df}{d\boldsymbol{\beta}} = \mathbf{0}$ とおくと，

$$\frac{df}{d\boldsymbol{\beta}} = \mathbf{0}$$
$$(\boldsymbol{n}' - \boldsymbol{m}')X = \mathbf{0}$$
$$\therefore \hat{\boldsymbol{m}}'X = \boldsymbol{n}'X$$

が成り立っている．

また，$f(\boldsymbol{\beta})$ の二階微分関数は，

$$\begin{aligned}\frac{d^2 f(\boldsymbol{\beta})}{d\boldsymbol{\beta}^2} &= \frac{d}{d\boldsymbol{\beta}}((\boldsymbol{n}' - \boldsymbol{m}')X) \\ &= \frac{d}{d\boldsymbol{\beta}}(-X'\boldsymbol{m}) \\ &= -X'\frac{d}{d\boldsymbol{\beta}}\boldsymbol{m}\end{aligned}$$

となる．

ここで，

$$X = \begin{bmatrix}\boldsymbol{x}_1' \\ \vdots \\ \boldsymbol{x}_q'\end{bmatrix} = [x_{ij}] \quad (i = 1, \ldots, q,\ j = 1, \ldots, r)$$

としておくと，$\log(\boldsymbol{m}) = X\boldsymbol{\beta}$ より，

$$\boldsymbol{m} = (m_1, \ldots, m_q)' = (\exp\{\boldsymbol{x}_1'\boldsymbol{\beta}\}, \ldots, \exp\{\boldsymbol{x}_q'\boldsymbol{\beta}\})'$$

と書ける．よって，

$$\begin{aligned}\frac{d}{d\boldsymbol{\beta}}\boldsymbol{m} &= \begin{bmatrix} x_{11}\exp\{\boldsymbol{x}_1'\boldsymbol{\beta}\} & \cdots & x_{1r}\exp\{\boldsymbol{x}_1'\boldsymbol{\beta}\} \\ \vdots & & \vdots \\ x_{q1}\exp\{\boldsymbol{x}_q'\boldsymbol{\beta}\} & \cdots & x_{qr}\exp\{\boldsymbol{x}_q'\boldsymbol{\beta}\} \end{bmatrix} \\ &= \begin{bmatrix} x_{11}m_1 & \cdots & x_{1r}m_1 \\ \vdots & & \vdots \\ x_{q1}m_q & \cdots & x_{qr}m_q \end{bmatrix} \\ &= \begin{bmatrix} m_1 & & O \\ & \ddots & \\ O & & m_q \end{bmatrix} X\end{aligned}$$

となる．これより，

$$\frac{d^2 f(\boldsymbol{\beta})}{d\boldsymbol{\beta}^2} = -X' \begin{bmatrix} m_1 & & O \\ & \ddots & \\ O & & m_q \end{bmatrix} X$$

である．
$f(\boldsymbol{\beta})$ の二階微分関数は負であることから，もとの関数 $L(\boldsymbol{m})$ を最大にする $\hat{\boldsymbol{m}}$ が存在するとき，それは一意的である． □

命題より，多項サンプリングモデルの期待度数の MLE $\hat{\boldsymbol{m}}$ には，

$$\hat{\boldsymbol{m}}' X = \boldsymbol{n}' X$$

が成り立つ．

検定

$\log(\boldsymbol{m}) = X\boldsymbol{\beta}$ と書ける対数線形モデルのことを，モデル X とよぶことにする．
検定ついて記述する前に，道具として有効である行列 X の列空間の定義を与える．

定義 3.1.1. 行列 X に対する

$$C(X) = \{\boldsymbol{v} \mid \boldsymbol{v} = X\boldsymbol{w} \text{ for some } \boldsymbol{w}\}$$

のことを X の列空間という．

この列空間 C により，あるモデル $\log(\boldsymbol{m}) = X\boldsymbol{\beta}$ と，$\boldsymbol{\beta}$ の要素のうち極大な効果項以外を排除し再構築したモデル $\log(\boldsymbol{m}) = X_0\boldsymbol{\beta}_0$ は同値であることが示される．

例 3.1.2. 2×2 分割表の飽和モデル

$$\log(m_{ij}) = u + u_{1(i)} + u_{2(j)} + u_{12(ij)}$$

について見てみる．このモデルは本章のはじめで記述したように，

$$\begin{bmatrix} \log(m_{11}) \\ \log(m_{12}) \\ \log(m_{21}) \\ \log(m_{22}) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ u_{1(1)} \\ u_{1(2)} \\ u_{2(1)} \\ u_{2(2)} \\ u_{12(11)} \\ u_{12(12)} \\ u_{12(21)} \\ u_{12(22)} \end{bmatrix}$$

と書ける．これを $\log(\boldsymbol{m}) = X\boldsymbol{\beta}$ としておく．
このモデルを再構築した極大項のみモデル $\log(m_{ij}) = u_{12(ij)}$ は，

$$\begin{bmatrix} \log(m_{11}) \\ \log(m_{12}) \\ \log(m_{21}) \\ \log(m_{22}) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_{12(11)} \\ u_{12(12)} \\ u_{12(21)} \\ u_{12(22)} \end{bmatrix}$$

と書ける．これを $\log(\boldsymbol{m}) = X_0\boldsymbol{\beta}_0$ と書く．
行列 X_0 は行列 X の列ベクトルからなる行列であるため，$C(X) = C(X_0)$ がいえる．つまり，モデル X で表されるデータは，モデル X_0 でも表されるということである．

また，列空間 C によりモデルの相対的な大小関係を表現できる．
モデル X_1 と X_2 の存在と $C(X_1) \subset C(X_2)$ の関係を仮定する．データが モデル X_1 で表される，つまり

$$\log(\boldsymbol{m}) = X_1\boldsymbol{\beta}_1$$

となる $\boldsymbol{\beta}_1$ が存在するとき，

$$\log(\boldsymbol{m}) = X_2\boldsymbol{\beta}_2$$

となる $\boldsymbol{\beta}_2$ も存在することになる．このような場合，モデル X_2 はモデル X_1 より相対的に大きいモデルといえる．

例 3.1.3. $2 \times 2 \times 2$ 表に対して，2 つのモデル

$$\text{モデル } X_1 : \log(m_{ijk}) = u_{12(ij)} + u_{3(k)}$$
$$\text{モデル } X_2 : \log(m_{ijk}) = u_{123(ijk)}$$

を仮定する．それぞれ

$$\text{モデル } X_1 : \begin{bmatrix} \log(m_{111}) \\ \log(m_{112}) \\ \log(m_{121}) \\ \log(m_{122}) \\ \log(m_{211}) \\ \log(m_{212}) \\ \log(m_{221}) \\ \log(m_{222}) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_{12(11)} \\ u_{12(12)} \\ u_{12(21)} \\ u_{12(22)} \\ u_{3(1)} \\ u_{3(2)} \end{bmatrix}$$

$$\text{モデル } X_2 : \begin{bmatrix} \log(m_{111}) \\ \log(m_{112}) \\ \log(m_{121}) \\ \log(m_{122}) \\ \log(m_{211}) \\ \log(m_{212}) \\ \log(m_{221}) \\ \log(m_{222}) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_{123(111)} \\ u_{123(112)} \\ u_{123(121)} \\ u_{123(122)} \\ u_{123(211)} \\ u_{123(212)} \\ u_{123(221)} \\ u_{123(222)} \end{bmatrix}$$

と書ける．このとき，$C(X_1) \subset C(X_2)$ が成り立つ．逆は成り立たない．これより，モデル X_1 で表されるデータは，モデル X_2 でも表すことができる．つまり，モデル X_2 はモデル X_1 と比べ相対的に大きい．

また，２つの例でみたように，任意の z 元分割表の z 因子交互作用モデル（飽和モデル）の計画行列は，単位行列 I_q となる．

検定について記述する．ここでの検定とは，ある分割表データに関して,「当てはまりがいいモデル X」に対する「X より小さいモデル X_0」の検定である．言い換えると，あるデータに関して $\log(\boldsymbol{m}) = X\boldsymbol{\beta}$ が成り立つとき,

$$\begin{array}{ll} \mathrm{H}_0 & : \log(\boldsymbol{m}) = X_0\boldsymbol{\beta}_0 \\ \mathrm{H}_1 & : \log(\boldsymbol{m}) = X\boldsymbol{\beta} \end{array} \qquad C(X_0) \subset C(X)$$

という「モデル X をさらに縮小できるかどうか」の検定である．このときの尤度比検定統計量は,

$$G^2 = -2\left[\log \frac{L(\hat{\boldsymbol{m}}_0)}{L(\hat{\boldsymbol{m}})}\right]$$

である．ここでの $\hat{\boldsymbol{m}}$ は，モデル X のもとでの $\boldsymbol{m}$ の MLE，$\hat{\boldsymbol{m}}_0$ はモデル X_0 のもとでの $\boldsymbol{m}$ の MLE のことである．この G^2 を変形すると,

$$\begin{aligned} G^2 &= -2[\ell(\hat{\boldsymbol{m}}_0) - \ell(\hat{\boldsymbol{m}})] \\ &= -2[\boldsymbol{n}' \log(\hat{\boldsymbol{m}}_0) - \boldsymbol{n}' \log(\hat{\boldsymbol{m}})] \\ &= 2\boldsymbol{n}'[\log(\hat{\boldsymbol{m}}) - \log(\hat{\boldsymbol{m}}_0)] \end{aligned}$$

となる．ここで

$$\begin{cases} \log(\hat{\boldsymbol{m}}) = X\hat{\boldsymbol{\beta}} \\ \log(\hat{\boldsymbol{m}}_0) = X_0\hat{\boldsymbol{\beta}}_0 \end{cases}, \ C(X_0) \subset C(X)$$

であるため,

$$\log(\hat{\boldsymbol{m}}_0) = X_0\hat{\boldsymbol{\beta}}_0 = X\hat{\boldsymbol{\gamma}}$$

となる $\hat{\boldsymbol{\gamma}}$ も存在する．

以上より G^2 は,

$$\begin{aligned} G^2 &= 2\boldsymbol{n}'[\log(\hat{\boldsymbol{m}}) - \log(\hat{\boldsymbol{m}}_0)] \\ &= 2\boldsymbol{n}'[X\hat{\boldsymbol{\beta}} - X\hat{\boldsymbol{\gamma}}] \\ &= 2\hat{\boldsymbol{m}}'X[\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\gamma}}] \quad (\because \text{MLE 条件} : \boldsymbol{n}'X = \hat{\boldsymbol{m}}'X) \\ &= 2\hat{\boldsymbol{m}}'[\log(\hat{\boldsymbol{m}}) - \log(\hat{\boldsymbol{m}}_0)] \\ &= 2\sum_{i=1}^{q} \hat{m}_i \log\left(\frac{\hat{m}_i}{\hat{m}_{0i}}\right) \end{aligned}$$

という 2 元・3 元分割表のときと同じ形式になることがわかる.

積多項サンプリング

t 個の多項母集団が存在し，全セル数が q の積多項サンプリングを仮定する．いくつかの表記を用意しておく.

- s_i : 第 i 母集団の多項カテゴリ数 $(i = 1, \ldots, t)$ $(\sum_i s_i = q)$
- n_{ij} : 第 i 母集団内の第 j カテゴリに該当する観測度数
  $(i = 1, \ldots, t,\ j = 1, \ldots, s_i)$
  (確率 p_{ij}, 期待度数 m_{ij} も同様に定める) $(\sum_j p_{ij} = 1,\ m_{ij} = n_{i\cdot}p_{ij})$
- $\boldsymbol{n} = (n_{11}, \ldots, n_{1s_1}, \ldots, n_{t1}, \ldots, n_{ts_t})'$: 観測度数ベクトル （q 次列ベクトル)
  (確率ベクトル $\boldsymbol{p}$, 期待度数ベクトル $\boldsymbol{m}$ も同様に定める)
- Z : 表示行列（$q \times t$ 行列）

ここでの表示行列 Z は，行が第 1 セルから第 q セルに対応，列が第 1 母集団から第 t 母集団に対応している．各母集団（列）とも，その母集団が保有するセル（行）に 1, それ以外には 0 が成分となる．1 が s 個並ぶベクトルを $\boldsymbol{J}_s = (1, \ldots, 1)'$ と書くと,

$$Z = \begin{bmatrix} \boldsymbol{J}_{s_1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J}_{s_2} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{J}_{s_1} \end{bmatrix} \quad (q \times t \text{ 行列})$$

である．この表示行列 Z により，各多項母集団の度数和と期待度数和は等しいという条件 $m_{i\cdot} = n_{i\cdot}$ は，$\boldsymbol{m}'Z = \boldsymbol{n}'Z$ と書ける.

例 3.1.4. 行和が固定の 2×3 分割表を与える.

	j	1	2	3	計
i	1	n_{11}	n_{12}	n_{13}	$n_{1\cdot}$
	2	n_{21}	n_{22}	n_{23}	$n_{2\cdot}$

このとき，$\boldsymbol{n} = (n_{11}, n_{12}, n_{13}, n_{21}, n_{22}, n_{23})'$ とすれば,

$$Z = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

と書ける.

積多項表に対し，対数線形モデル $\log(\boldsymbol{m}) = X\boldsymbol{\beta}$ を仮定する.

期待度数の MLE $\hat{\boldsymbol{m}}$ について考える．多項モデルと同様，尤度関数

$$L(\boldsymbol{p}) = \prod_{i=1}^{t} \left[\frac{n_{i\cdot}!}{\prod_{j=1}^{s_i} n_{ij}!} \prod_{j=1}^{s_i} p_{ij}^{n_{ij}} \right]$$

に $p_{ij} = m_{ij}/n_{i\cdot}$ を代入する.

$$L(\boldsymbol{m}) = \prod_{i=1}^{t} \left[\frac{n_{i\cdot}!}{\prod_{j=1}^{s_i} n_{ij}!} \prod_{j=1}^{s_i} \left(\frac{m_{ij}}{n_{i\cdot}} \right)^{n_{ij}} \right]$$

これを対数変換し，$\boldsymbol{m}$ に依存する項のみ残すと,

$$\ell(\boldsymbol{m}) = \sum_{i=1}^{t} \sum_{j=1}^{s_i} n_{ij} \log(m_{ij}) = \boldsymbol{n}' \log(\boldsymbol{m})$$

となる．これを最大にする $\boldsymbol{m}$=$\hat{\boldsymbol{m}}$ が MLE である.

$\hat{\boldsymbol{m}}$ には次の 2 つの制約

- $\log(\hat{\boldsymbol{m}}) = X\hat{\boldsymbol{\beta}}$ ：対数線形構造
- $\hat{\boldsymbol{m}}'Z = \boldsymbol{n}'Z$ ：母集団度数和と期待度数和の一致

があり，また多項表と同様,

$$\hat{\boldsymbol{m}}'X = \boldsymbol{n}'X$$

を満たす（証明は命題 3.1.1 と同様である）.

3.2 漸近理論

本節では，標本数が大きい場合に成立する漸近分布について簡潔に記述する．詳しい証明は，参考文献 [5]，[6] などを参照されたい.

多項サンプリング

多項表のもとでの対数線形モデル $\log(\boldsymbol{m}) = X\boldsymbol{\beta}$ に関する漸近分布等について述べる．事前にいくつかの表記を用意しておく．
任意のベクトル $\boldsymbol{v} = (v_1, \ldots, v_q)'$ に対して，

$$D(\boldsymbol{v}) = [d_{ij}],\ d_{ij} = \begin{cases} v_i & (i = j) \\ 0 & (i \neq j) \end{cases} \quad i = 1, \ldots, q$$

のように対角行列 $D(\boldsymbol{v})$ を定める．特に，これに確率ベクトル $\boldsymbol{p}$ を適用したものを $D = D(\boldsymbol{p})$ と定める．また，

- $\hat{\boldsymbol{\mu}}$: $\boldsymbol{\mu}$ の MLE （$\log(\hat{\boldsymbol{m}})$）
- $\boldsymbol{J} = (1, \ldots, 1)'$ （q 次ベクトル）
- $A = X(X'DX)^{-1}X'D$
- $A_Z = \boldsymbol{J}(\boldsymbol{J}'D\boldsymbol{J})^{-1}\boldsymbol{J}'D$

と定める．また仮定として，モデル計画行列 X は $(X'DX)^{-1}$ が存在するように選ばれているとする．

ここで，多項サンプリングの場合，$D(\boldsymbol{m}) = n.D$ であるから，A と A_Z の定義内の D の代わりに $D(\boldsymbol{m})$ を利用することもできる．

- $A = X(X'D(\boldsymbol{m})X)^{-1}X'D(\boldsymbol{m})$
- $A_Z = \boldsymbol{J}(\boldsymbol{J}'D(\boldsymbol{m})\boldsymbol{J})^{-1}\boldsymbol{J}'D(\boldsymbol{m})$

$\boldsymbol{\mu}$ の MLE $\hat{\boldsymbol{\mu}}$ に関する次の定理を与える．

定理 3.2.1. [6, p.323]
総度数 $n.$ の多項サンプリング表の対数線形モデル $\boldsymbol{\mu} = X\boldsymbol{\beta}$ について，次が成立する．

(a) $n.$ が十分に大きいとき，$\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}$ は漸近分布

$$N(\boldsymbol{0}, [A - A_Z]D^{-1}(\boldsymbol{m}))$$

に従う．

(b) $n.$ が大きくなるにつれ，$\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}$ は $\boldsymbol{0}$ に確率収束する．

(c) $n.$ が十分に大きいとき，$\hat{\boldsymbol{m}} - \boldsymbol{m}$ は漸近分布

$$N(\boldsymbol{0}, D(\boldsymbol{m})[A - A_Z])$$

に従う．

(d) $n.$ が大きくなるにつれ，$\dfrac{\hat{\boldsymbol{m}}}{n.}$ は $\boldsymbol{p}$ に確率収束する．

系 3.2.1. [6, p.324]
$(X'DX)^{-1}$ が存在し，$\hat{\boldsymbol{\beta}}$ が $\hat{\boldsymbol{\mu}} = X\hat{\boldsymbol{\beta}}$ を満たすならば，$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ は $\mathbf{0}$ に確率収束する．

この漸近分布に関する定理を用いて，第 i セルに関するパラメータ p_i, m_i, μ_i についての推論ができる．パラメータベクトル $\boldsymbol{p}$,$\boldsymbol{m}$,$\boldsymbol{\mu}$ には

$$\begin{aligned}
\boldsymbol{m} &= (m_1, \ldots, m_q)' \\
\boldsymbol{\mu} &= \log(\boldsymbol{m}) = (\log(m_1), \ldots, \log(m_q))' \\
\boldsymbol{p} &= \frac{1}{n.}\boldsymbol{m} = \left(\frac{m_1}{n.}, \ldots, \frac{m_q}{n.}\right)'
\end{aligned}$$

という関係がある．
その推論に必要な多変量正規分布に関する定理を記述しておく．

定理 3.2.2. n 次確率変数ベクトル $\boldsymbol{X}$ が多変量正規分布 $N(\boldsymbol{\mu}, \Sigma)$ に従うとき，任意の n 次ベクトル $\boldsymbol{\rho}$ に対して

$$\boldsymbol{\rho}'\boldsymbol{X} \sim N(\boldsymbol{\rho}'\boldsymbol{\mu}, \boldsymbol{\rho}'\Sigma\boldsymbol{\rho})$$

が成り立つ．

（証明）
$Y \sim N(\mu, \sigma^2)$ の積率母関数 $M_Y(t)$ は $\exp\left\{\mu t + \dfrac{\sigma^2 t^2}{2}\right\}$ である．
$\boldsymbol{Z} \sim N(\mathbf{0}, I_n)$ とすると，その $\boldsymbol{Z}$ の積率母関数 M は，

$$\begin{aligned}
M_{\boldsymbol{Z}}(\boldsymbol{t}) &= \mathrm{E}[\exp \boldsymbol{t}'\boldsymbol{Z}] \\
&= \prod_{i=1}^{n} \mathrm{E}[\exp\{t_i Z_i\}] \\
&= \prod_{i=1}^{n} M_{Z_i}(t_i) \qquad (\because M_{Z_i}(t_i) : Z_i\text{の積率母関数}) \\
&= \prod_{i=1}^{n} \exp\left\{\frac{1}{2}t_i^2\right\} \qquad (\because Z_i \sim N(0,1)\text{ の積率母関数}) \\
&= \exp\left\{\sum_{i=1}^{n} \frac{1}{2}t_i^2\right\} \\
&= \exp\left\{\frac{1}{2}\boldsymbol{t}'\boldsymbol{t}\right\}
\end{aligned}$$

と書ける．
$\boldsymbol{X} = \boldsymbol{\mu} + \Sigma^{\frac{1}{2}}\boldsymbol{Z}$ は $N(\boldsymbol{\mu}, \Sigma)$ に従う．その $\boldsymbol{X}$ の積率母関数 $M_{\boldsymbol{X}}$ は，

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = \mathrm{E}[\exp\{\boldsymbol{t}'\boldsymbol{X}\}]$$

$$= \mathrm{E}[\exp\{\boldsymbol{t}'\boldsymbol{\mu} + \boldsymbol{t}'\Sigma^{\frac{1}{2}}\boldsymbol{Z}\}]$$
$$= \exp\{\boldsymbol{t}'\boldsymbol{\mu}\}\mathrm{E}[\exp\{\boldsymbol{t}'\Sigma^{\frac{1}{2}}\boldsymbol{Z}\}]$$

と変形できる．
ここで，$\mathrm{E}[\exp\{\boldsymbol{t}'\Sigma^{\frac{1}{2}}\boldsymbol{Z}\}]$ について，$\boldsymbol{t}'\Sigma^{\frac{1}{2}} = \boldsymbol{u}'$ とおくと，

$$\begin{aligned}
\mathrm{E}[\exp\{\boldsymbol{t}'\Sigma^{\frac{1}{2}}\boldsymbol{Z}\}] &= \mathrm{E}[\exp\{\boldsymbol{u}'\boldsymbol{Z}\}] \\
&= \exp\left\{\frac{1}{2}\boldsymbol{u}'\boldsymbol{u}\right\} \quad (\because \boldsymbol{Z}\text{の積率母関数}) \\
&= \exp\left\{\frac{1}{2}\boldsymbol{t}'\Sigma^{\frac{1}{2}}\left(\Sigma^{\frac{1}{2}}\right)'\boldsymbol{t}\right\} \\
&= \exp\left\{\frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}\right\}
\end{aligned}$$

と書ける．この結果を用いると，

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = \exp\left\{\boldsymbol{t}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}\right\}$$

と書ける．

ここで $\boldsymbol{\rho}'\boldsymbol{X}$ の積率母関数 $M_{\boldsymbol{\rho}'\boldsymbol{X}}$ は，

$$\begin{aligned}
M_{\boldsymbol{\rho}'\boldsymbol{X}}(t) &= \mathrm{E}[\exp\{t\boldsymbol{\rho}'\boldsymbol{X}\}] \\
&= M_{\boldsymbol{X}}(t\boldsymbol{\rho}) \\
&= \exp\left\{(t\boldsymbol{\rho})'\boldsymbol{\mu} + \frac{1}{2}(t\boldsymbol{\rho})'\Sigma(t\boldsymbol{\rho})\right\} \quad (\because \boldsymbol{X}\text{の積率母関数}) \\
&= \exp\left\{t(\boldsymbol{\rho}'\boldsymbol{\mu}) + \frac{1}{2}t^2\boldsymbol{\rho}'\Sigma\boldsymbol{\rho}\right\}
\end{aligned}$$

と書けるが，これは $N(\boldsymbol{\rho}'\boldsymbol{\mu}, \boldsymbol{\rho}'\Sigma\boldsymbol{\rho})$ の積率母関数でもある．
つまり，$\boldsymbol{\rho}'\boldsymbol{X} \sim N(\boldsymbol{\rho}'\boldsymbol{\mu}, \boldsymbol{\rho}'\Sigma\boldsymbol{\rho})$ である． □

q 次のベクトルで，第 i 成分が 1 で他は 0 のベクトル $\boldsymbol{e}_i'$ を用意する．

$$\boldsymbol{e}_i' = (0, \ldots, 0, 1, 0, \ldots, 0)$$

この $\boldsymbol{e}_i'$ を用いると，

$$\begin{cases}
\hat{\mu}_i - \mu_i &= \boldsymbol{e}_i'(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \\
\hat{m}_i - m_i &= \boldsymbol{e}_i'(\hat{\boldsymbol{m}} - \boldsymbol{m}) \\
\hat{p}_i - p_i &= \boldsymbol{e}_i'(\hat{\boldsymbol{p}} - \boldsymbol{p})
\end{cases}$$

と書ける．ここで定理 3.2.1,3.2.2 より，

$$\begin{cases}
\hat{\mu}_i - \mu_i &\sim N(0, \boldsymbol{e}_i'[A - A_Z]D^{-1}(\boldsymbol{m})\boldsymbol{e}_i) \\
\hat{m}_i - m_i &\sim N(0, \boldsymbol{e}_i'D(\boldsymbol{m})[A - A_Z]\boldsymbol{e}_i) \\
\hat{p}_i - p_i &\sim N\left(0, \frac{1}{n_\cdot^2}\boldsymbol{e}_i'D(\boldsymbol{m})[A - A_Z]\boldsymbol{e}_i\right)
\end{cases}$$

という漸近分布が得られる．各分布の分散を整理すると，

$$\begin{cases} \hat{\mu}_i - \mu_i & \sim & N\left(0, \dfrac{a_{ii}}{m_i} - \dfrac{1}{n.}\right) \\ \hat{m}_i - m_i & \sim & N\left(0, m_i a_{ii} - \dfrac{m_i^2}{n.}\right) \\ \hat{p}_i - p_i & \sim & N\left(0, \dfrac{p_i a_{ii}}{n.} - \dfrac{p_i^2}{n.}\right) \end{cases}$$

となる．a_{ii} は先に定められた行列 A の第 (i,i) 成分である．μ_i, m_i, p_i の信頼区間の算出や検定は，推定された $\hat{m}_i, \hat{p}_i$ を標準化分布

$$\begin{cases} \dfrac{\hat{\mu}_i - \mu_i}{\sqrt{\dfrac{\hat{a}_{ii}}{\hat{m}_i} - \dfrac{1}{n.}}} & \sim & N(0,1) \\ \dfrac{\hat{m}_i - m_i}{\sqrt{\hat{m}_i \hat{a}_{ii} - \dfrac{\hat{m}_i^2}{n.}}} & \sim & N(0,1) \\ \dfrac{\hat{p}_i - p_i}{\sqrt{\dfrac{\hat{p}_i \hat{a}_{ii}}{n.} - \dfrac{\hat{p}_i^2}{n.}}} & \sim & N(0,1) \end{cases}$$

に基づいて行う．

積多項サンプリング

多項サンプリングとは少し異なるが，同様な漸近結果が積多項サンプリングにも存在する．$\log(\boldsymbol{m}) = X\boldsymbol{\beta}$ を仮定し，いくつか道具を再定義する．

- $\boldsymbol{m}^* = (m^*_{11}, \ldots, m^*_{ts_t})'$, $\left(m^*_{ij} = \frac{n_{i\cdot} p_{ij}}{n_{\cdot\cdot}}\right)$　(q 次列ベクトル)
- $D = D(\boldsymbol{m}^*)$　($q \times q$ 行列)
- $A = X(X'DX)^{-1}X'D$　($q \times q$ 行列)

  ただし，変則的ではあるが，

$$A = \begin{bmatrix} a_{11,11} & a_{12} & \cdots & & a_{1q} \\ a_{21} & a_{12,12} & & & \\ & & \ddots & & \\ \vdots & & a_{1s_1,1s_1} & & \vdots \\ & & & \ddots & \\ & & & a_{t1,t1} & \\ & & & & \ddots \\ a_{q1} & & & & a_{ts_t,ts_t} \end{bmatrix}$$

  と表すことにする．

- $A_Z = Z(Z'DZ)^{-1}Z'D \quad (q \times q \text{ 行列})$

そもそも積多項分布とは，その名の通り多項分布の積の分布であった．これより，多項サンプリングは，積多項サンプリングの特別な場合であると考えられる．実際，

- $t = 1$ (多項母集団数)
- $Z = (1, \ldots, 1)'$ (q 次列ベクトル)
- $n_{1\cdot} = n_{\cdot\cdot}$
- $\boldsymbol{m}^* = \boldsymbol{p} \quad \left(\because\ m^*_{1j} = \frac{n_{1\cdot}p_{1j}}{n_{\cdot\cdot}} = \frac{n_{1\cdot}p_{1j}}{n_{1\cdot}}\right)$

とおけば，多項サンプリングは積多項サンプリング特殊な場合であることがわかる．

積多項サンプリングについての漸近結果を記述する．これは多項サンプリングにも適用可能である．

定理 3.2.3. [6, p.341]
多項母集団が t 個存在する積多項サンプリング表に対し，$\boldsymbol{\mu} = \log(\boldsymbol{m}) = X\boldsymbol{\beta}$ を仮定する．各母集団の度数和 $n_{1\cdot}, \ldots, n_{t\cdot}$ が十分大きいとき次が成立する．

(a) $\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}$ は漸近分布

$$N(\mathbf{0}, [A - A_Z]D^{-1}(\boldsymbol{m}))$$

に従う．

(b) $\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}$ は $\mathbf{0}$ に確率収束する．

(c) $\hat{\boldsymbol{m}} - \boldsymbol{m}$ は漸近分布

$$N(\mathbf{0}, D(\boldsymbol{m})[A - A_Z])$$

に従う．

(b) $\dfrac{\hat{\boldsymbol{m}}}{n_{\cdot\cdot}}$ は $\boldsymbol{m}^*$ に確率収束する．

この定理により，多項サンプリングのときと同様，個々のパラメータに関する分布が得られる．

$$\begin{cases} \dfrac{\hat{\mu}_{ij} - \mu_{ij}}{\sqrt{\dfrac{\hat{a}_{ij,ij}}{\hat{m}_{ij}} - \dfrac{1}{n_{i\cdot}}}} & \sim \quad N(0,1) \\ \dfrac{\hat{m}_{ij} - m_{ij}}{\sqrt{\hat{m}_{ij}\hat{a}_{ij,ij} - \dfrac{\hat{m}^2_{ij}}{n_{i\cdot}}}} & \sim \quad N(0,1) \\ \dfrac{\hat{p}_{ij} - p_{ij}}{\sqrt{\dfrac{\hat{p}_{ij}\hat{a}_{ij,ij}}{n_{i\cdot}} - \dfrac{\hat{p}^2_{ij}}{n_{i\cdot}}}} & \sim \quad N(0,1) \end{cases}$$

最後に，大標本のもとでの積多項サンプリングモデルの仮説検定に関する定理を挙げる．

定理 3.2.4. [6, p.342]
対数線形モデル $\boldsymbol{\mu} = X\boldsymbol{\beta}$ を仮定する．ある行列 X_0 は $C(X_0) \subset C(X)$ を満たし，また，$\mathrm{rank}(X_0) = \gamma_0, \mathrm{rank}(X) = \gamma$ であるとする．ここで，次の検定

$$\begin{aligned} \mathrm{H}_0 : &\quad \boldsymbol{\mu} = X_0\boldsymbol{\beta}_0 \quad \text{for some } \boldsymbol{\beta}_0 \\ \mathrm{H}_1 : &\quad \boldsymbol{\mu} \neq X_0\boldsymbol{\beta}_0 \quad \text{for any } \boldsymbol{\beta}_0 \end{aligned}$$

に関して，$n_{..}$ が十分大のとき，次の漸近分布が成り立つ．

(a) H_0 が真ならば，$G^2 \sim \chi^2_{\gamma-\gamma_0}$

(b) H_0 が真ならば，$X^2 \sim \chi^2_{\gamma-\gamma_0}$

(c) H_0 が真ならば，$G^2 - X^2$ は 0 に確率収束する．

(d) H_0 が偽ならば，G^2 と X^2 は $n_{..}$ 大きくなるにつれ無限に近づく．

例 3.2.1. 3×3 分割表について，モデル $\log(m_{ij}) = u_{1(i)} + u_{2(j)}$ の対飽和モデル $\log(m_{ij}) = u_{12(ij)}$ の検定を行いたい．モデル $\log(m_{ij}) = u_{1(i)} + u_{2(j)}$ の計画行列 X_0 は，

$$X_0 = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

と書ける（$\boldsymbol{\beta} = (u_{1(1)}, u_{1(2)}, u_{1(3)}, u_{2(1)}, u_{2(2)}, u_{2(3)})'$）．このとき，$\mathrm{rank}(X_0) = 5$ である．飽和モデルの計画行列 X は，9 次単位行列 I_9 であり，$\mathrm{rank}(X) = 9$ である．これより，モデル X_0 の飽和モデル X に対する検定の検定統計量 G^2 については，

$$G^2 \sim \chi^2_4$$

となる．既に見たように，この自由度”4”は，２つのモデルの自由なパラメータの数の差である．

3.3 MLE 算出法

3.3.1 ニュートン法を用いた $\hat{m}$ の算出

ニュートン法を用いた $\hat{\boldsymbol{m}}$ の算出法を記述する．
先に，ベクトル方程式 $\boldsymbol{f}(\boldsymbol{\beta}) = \mathbf{0}$ の解 $\boldsymbol{\beta}$ をニュートン法で算出する方法を簡単にまとめておく．

1. $\boldsymbol{\beta}$ の初期値 $\boldsymbol{\beta}_0$ をとる

2. $\boldsymbol{df}(\boldsymbol{\beta}_t)$ を，$\boldsymbol{\beta}_t$ に関する $\boldsymbol{f}(\boldsymbol{\beta}_t)$ の偏導関数行列とし，次の式

$$\boldsymbol{\beta}_{t+1} = \boldsymbol{\beta}_t - [\boldsymbol{df}(\boldsymbol{\beta}_t)]^{-1}\boldsymbol{f}(\boldsymbol{\beta}_t)$$

   の計算を繰り返す.

3. 2 の計算を繰り返した結果，ある $\boldsymbol{\beta}$ に収束した場合，それを解とする.

この方法を対数線形モデルの $\boldsymbol{m}$ の MLE 算出に適用する.

第 3.1 節の $\log(\boldsymbol{v})$ と同様の形式的な表記として，ベクトル $\boldsymbol{v}= (v_1, \ldots, v_q)'$ に対して

- $\exp(\boldsymbol{v}) = (\exp v_1, \ldots, \exp v_q)'$

と定めておく.

対数線形モデル $\log(\boldsymbol{m}) = X\boldsymbol{\beta}$ の $\boldsymbol{m}$ の MLE というのは，対数尤度の一部の項である

$$\ell(\boldsymbol{m}) = \boldsymbol{n}' \log(\boldsymbol{m})$$

の $\boldsymbol{m}$ に関する偏導関数が 0 になる点 $\hat{\boldsymbol{m}}$ である.

$\log(\boldsymbol{m}) = X\boldsymbol{\beta}$ より，$\boldsymbol{m}(\boldsymbol{\beta}) = \exp(X\boldsymbol{\beta})$ のように $\boldsymbol{m}$ は $\boldsymbol{\beta}$ の関数として書ける．これより

$$\ell(\boldsymbol{m}) = \ell(\exp\{X\boldsymbol{\beta}\})$$

として書ける.

この ℓ の $\boldsymbol{\beta}$ に関する偏導関数行列 $\boldsymbol{d\ell}(\exp\{X\boldsymbol{\beta}\})$ を $\boldsymbol{f}(\boldsymbol{\beta})$ とおくと，

$$\boldsymbol{f}(\boldsymbol{\beta}) = \boldsymbol{0}$$

の解 $\hat{\boldsymbol{\beta}}$ はニュートン法で算出され，それを用いた $\hat{\boldsymbol{m}} = \exp(X\hat{\boldsymbol{\beta}})$ が $\boldsymbol{m}$ の MLE となる.

命題 3.1.1 の証明における結果を引用すると，

$$\begin{cases} \boldsymbol{f}(\boldsymbol{\beta}_t) = X'(\boldsymbol{n} - \boldsymbol{m}(\boldsymbol{\beta}_t)) \\ \boldsymbol{df}(\boldsymbol{\beta}_t) = -X'D(\boldsymbol{m}(\boldsymbol{\beta}_t))X \end{cases}$$

である．これより，

$$\boldsymbol{\beta}_{t+1} = \boldsymbol{\beta}_t + [X'D(\boldsymbol{m}(\boldsymbol{\beta}_t))X]^{-1}X'(\boldsymbol{n} - \boldsymbol{m}(\boldsymbol{\beta}_t))$$

がニュートン法の漸化式となり，ある $\hat{\boldsymbol{\beta}}$ に収束するまで計算を行い MLE を算出する.

3.3.2 比例反復法を用いた $\hat{m}$ の算出

比例反復法を用いた $\hat{\boldsymbol{m}}$ の算出を，3 元分割表のモデル

$$M^{(8)} : \frac{p_{111}p_{ij1}}{p_{i11}p_{1j1}} = \frac{p_{11k}p_{ijk}}{p_{i1k}p_{1jk}} \quad \text{for all } i = 2, \cdots, I,\ j = 2, \cdots, J,\ k = 2, \cdots, K$$

を例に適用してみる．

モデル $M^{(8)}$ の $\hat{m}_{ijk}$ には，

$$\hat{m}_{ij\cdot} = n_{ij\cdot},\ \hat{m}_{i\cdot k} = n_{i\cdot k},\ \hat{m}_{\cdot jk} = n_{\cdot jk}$$

という周辺制約があった．これより，

$$1 = \frac{n_{ij\cdot}}{\hat{m}_{ij\cdot}} = \frac{n_{i\cdot k}}{\hat{m}_{i\cdot k}} = \frac{n_{\cdot jk}}{\hat{m}_{\cdot jk}}$$

が成り立ち，これより

$$\begin{cases} \hat{m}_{ijk} = \dfrac{n_{ij\cdot}}{\hat{m}_{ij\cdot}}\hat{m}_{ijk} \\ \hat{m}_{ijk} = \dfrac{n_{i\cdot k}}{\hat{m}_{i\cdot k}}\hat{m}_{ijk} \\ \hat{m}_{ijk} = \dfrac{n_{\cdot jk}}{\hat{m}_{\cdot jk}}\hat{m}_{ijk} \end{cases} \tag{3.3}$$

も成り立つことになる．適切な初期値 $\hat{m}_{ijk}^{[0]}$ を与え，式 (3.3) を用いて繰り返し修正し，ある値 $\hat{m}_{ijk}$ に収束するまで続ける．具体的には，ある時点での推定値 $\hat{m}_{ijk}^{[t]}$ が与えられたとき，それを

$$\begin{aligned} \hat{m}_{ijk}^{[3t+1]} &= \frac{n_{ij\cdot}}{\hat{m}_{ij\cdot}^{[3t]}}\hat{m}_{ijk}^{[3t]} \\ \hat{m}_{ijk}^{[3t+2]} &= \frac{n_{i\cdot k}}{\hat{m}_{i\cdot k}^{[3t+1]}}\hat{m}_{ijk}^{[3t+1]} \\ \hat{m}_{ijk}^{[3(t+1)]} &= \frac{n_{\cdot jk}}{\hat{m}_{\cdot jk}^{[3t+2]}}\hat{m}_{ijk}^{[3t+2]} \end{aligned}$$

と修正する．修正を繰り返し，収束すなわち任意の i, j, k に対して

$$\hat{m}_{ijk}^{[3t]} \fallingdotseq \hat{m}_{ijk}^{[3t+1]} \fallingdotseq \hat{m}_{ijk}^{[3t+2]} \fallingdotseq \hat{m}_{ijk}^{[3(t+1)]}$$

となったとき修正を終了する．

なお，初期値は通常

$$\hat{m}_{ijk}^{[0]} = 1 \quad \text{for all } i, j, k$$

ととる．これは，期待度数 m に置き換えた任意のオッズ比が 1 になるためである．任意のオッズ比が 1 になるということは，第 1 章や第 2 章でみたような任意の標準的なモデルのオッズ比の条件を満たすということであるから，そのモデルに適した初期値であるといえる．

以上の比例反復法の方法を，一般的な表現で記述しておく．添字をまとめて $\boldsymbol{i}$ と表し，1 つ以上の添字について和をとった状態を $\boldsymbol{j}_{\cdot}$ と表すことにする．

1. $m_{\boldsymbol{i}}$ の初期推定値 $\hat{m}_{\boldsymbol{i}}^{[0]}$ を 1 とおく.

2. モデルの $m_{\boldsymbol{i}}$ に関する周辺制約から更新式を導く.

$$\hat{m}_{\boldsymbol{i}}^{[t+1]} = \frac{n_{\boldsymbol{j}\cdot}}{\hat{m}_{\boldsymbol{j}\cdot}^{[t]}} \hat{m}_{\boldsymbol{i}}^{[t]} \tag{3.4}$$

3. ある値 $\hat{m}_{\boldsymbol{i}}$ に収束するまで (3.4) の計算を繰り返す.

3.3.3 母数ベクトル $\hat{\beta}$ の算出

モデル $\log(\boldsymbol{m}) = X\boldsymbol{\beta}$ の $\boldsymbol{\beta}$ の MLE の算出について記述する．$\hat{\boldsymbol{\beta}}$ は，先に推定された $\hat{\boldsymbol{m}}$ を利用して算出する.

例 3.3.1. 2 元分割表のモデル

$$\log(m_{ij}) = u + u_{1(i)} + u_{2(j)} + u_{12(ij)}$$

の各パラメータに,

$$u_{1(\cdot)} = u_{2(\cdot)} = u_{12(i\cdot)} = u_{12(\cdot j)} = 0$$

という制約を仮定する．ここで推定された期待度数，$\log(\hat{m}_{ij}) = \omega_{ij}$ とおくと,

$$\begin{aligned} \omega_{\cdot\cdot} &= \log(\hat{m}_{\cdot\cdot}) \\ &= IJ\hat{u} + \hat{u}_{1(\cdot)} + \hat{u}_{2(\cdot)} + \hat{u}_{1(\cdot\cdot)} \\ &= IJ\hat{u} \\ \therefore \hat{u} &= \frac{1}{IJ}\omega_{\cdot\cdot} = \bar{\omega}_{\cdot\cdot} \end{aligned}$$

と書ける．同様に,

$$\begin{aligned} \omega_{i\cdot} &= J\hat{u} + J\hat{u}_{1(i)} + \hat{u}_{2(\cdot)} + \hat{u}_{1(i\cdot)} \\ \therefore \hat{u}_{1(i)} &= \frac{1}{J}\omega_{i\cdot} - \hat{u} = \bar{\omega}_{i\cdot} - \bar{\omega}_{\cdot\cdot} \end{aligned}$$

と書ける．残りについても同様に,

$$\begin{cases} \hat{u}_{2(j)} & = \bar{\omega}_{\cdot j} - \bar{\omega}_{\cdot\cdot}, \\ \hat{u}_{12(ij)} & = \omega_{ij} - \bar{\omega}_{i\cdot} - \bar{\omega}_{\cdot j} + \bar{\omega}_{\cdot\cdot} \end{cases}$$

と書ける.

この例のように，$\hat{\boldsymbol{\beta}}$ は $\hat{\boldsymbol{\mu}}(= \log(\hat{\boldsymbol{m}}))$ を利用して算出する．この値は，制約をもとに $\boldsymbol{\rho}$ を設定し，$\hat{\boldsymbol{\mu}}$ にかけた $\boldsymbol{\rho}$'$\hat{\boldsymbol{\mu}}$ でもある.

例 3.3.2. 前の例のモデルに戻る．分割表のサイズを 2×2 とし，$\hat{\boldsymbol{\mu}} = (\hat{\mu}_{11}, \hat{\mu}_{12}, \hat{\mu}_{21}, \hat{\mu}_{22})'$ とする．存在する各パラメータに関して次の表のように $\boldsymbol{\rho}$ を割り当てると，$\boldsymbol{\rho}$'$\hat{\boldsymbol{\mu}}$ でそのパラメータの推定値が算出される．

β_i	$\boldsymbol{\rho}$'		β_i	$\boldsymbol{\rho}$'
u	$\frac{1}{4}(1,1,1,1)$		$u_{12(11)}$	$\frac{1}{4}(1,-1,-1,1)$
$u_1(1)$	$\frac{1}{4}(1,1,-1,-1)$		$u_{12(12)}$	$\frac{1}{4}(-1,1,1,-1)$
$u_1(2)$	$\frac{1}{4}(-1,-1,1,1)$		$u_{12(21)}$	$\frac{1}{4}(-1,1,1,-1)$
$u_2(1)$	$\frac{1}{4}(1,-1,1,-1)$		$u_{12(22)}$	$\frac{1}{4}(1,-1,-1,1)$
$u_2(2)$	$\frac{1}{4}(-1,1,-1,1)$			

例えば，$u_{12(22)}$ なら，

$$\begin{aligned}
\boldsymbol{\rho}'\hat{\boldsymbol{\mu}} &= \frac{1}{4}(1,-1,-1,1)\hat{\boldsymbol{\mu}} \\
&= \frac{1}{4}(\hat{\mu}_{11} - \hat{\mu}_{12} - \hat{\mu}_{21} + \hat{\mu}_{22}) \\
&= \frac{1}{4}(\hat{u}_{12(11)} - \hat{u}_{12(12)} - \hat{u}_{12(21)} + \hat{u}_{12(22)}) \\
&= \frac{1}{4}(\hat{u}_{12(22)} + \hat{u}_{12(22)} + \hat{u}_{12(22)} + \hat{u}_{12(22)}) \quad (\because u_{12(i\cdot)} = u_{12(\cdot j)} = 0) \\
&= \hat{u}_{12(22)}
\end{aligned}$$

のように算出される．

例 3.3.3. 3×3 表の場合でも前と同じモデルを考える．
$\hat{\boldsymbol{\mu}} = (\hat{\mu}_{11}, \hat{\mu}_{12}, \hat{\mu}_{13}, \hat{\mu}_{21}, \hat{\mu}_{22}, \hat{\mu}_{23}, \hat{\mu}_{31}, \hat{\mu}_{32}, \hat{\mu}_{33})'$ とする．このとき，

$$\boldsymbol{\rho}' = \frac{1}{9}(2,2,2,-1,-1,-1,-1,-1,-1)$$

とおくと，

$$\begin{aligned}
\boldsymbol{\rho}'\hat{\boldsymbol{\mu}} &= \frac{1}{9}(2\hat{\mu}_{11} + 2\hat{\mu}_{12} + 2\hat{\mu}_{13} - \hat{\mu}_{21} - \hat{\mu}_{22} - \hat{\mu}_{23} - \hat{\mu}_{31} - \hat{\mu}_{32} - \hat{\mu}_{33}) \\
&= \frac{1}{9}\{2(\hat{\mu}_{11} + \hat{\mu}_{12} + \hat{\mu}_{13}) - 6\hat{u} - 3\hat{u}_{1(2)} - 3\hat{u}_{1(3)}\} \quad (\because u_{2(\cdot)} = u_{12(i\cdot)} = 0) \\
&= \frac{1}{9}\{2(3\hat{u} + 3\hat{u}_{1(1)}) - 6\hat{u} - 3(\hat{u}_{1(2)} + \hat{u}_{1(3)})\} \\
&= \frac{1}{9}\{6\hat{u}_{1(1)} - 3(-\hat{u}_{1(1)})\} \quad (\because u_{1(\cdot)} = 0) \\
&= \hat{u}_{1(1)}
\end{aligned}$$

のように算出される．

いまのように，母数の推定値は $\boldsymbol{\rho}$'$\hat{\boldsymbol{\mu}}$ で算出されることから，定理 3.2.3(a) より，次の漸近分布

$$\frac{\boldsymbol{\rho}'\hat{\boldsymbol{\mu}} - \boldsymbol{\rho}'X\boldsymbol{\beta}}{\sqrt{\boldsymbol{\rho}'(A - A_z)D^{-1}(\boldsymbol{n})\boldsymbol{\rho}}} \sim N(0,1)$$

が利用でき，検定等が可能である．

第4章　分割表モデル選択

4.1　モデル選択規準

分割表の対数線形モデルは，属性の数に応じていくつも候補が考えられる．それらのモデルの良さを比較する道具として，**モデル選択規準**がある．下に挙げるモデル選択規準にはもともとの一般的な形式が存在するが，本論文では，分割表解析について記述してきたため，これまでの表記・道具を用いた形式で与える．

決定係数 R^2

興味がある対数線形モデルのうち，最小のモデルを X_0 とし，G^2 を飽和モデルに対する尤度比検定統計量とすると，あるモデル X の決定係数 R^2 は

$$R^2 = \frac{G^2(X_0) - G^2(X)}{G^2(X_0)}$$

のように与えられる．変形により，$R^2 = 1 - G^2(X)/G^2(X_0)$ と書ける．$G^2(X)$ が小さい，すなわち X の当てはまりが良いときほど R^2 の値は大きくなる．最小モデル X_0 は固定であるから，R^2 の値で他のモデルと比較が可能である．この R^2 が大きいほど良いモデルと判断する．

自由度調整済決定係数 $Adjusted\ R^2$

R^2 は，モデルの母数の数が多い，すなわち大きいモデルほどモデルの当てはまりがよくなるため R^2 の値が大きくなってしまう欠点がある．母数が多すぎるものは，解析者にとっては良いモデルとは言えない．そこで，その欠点を解消するために R^2 を修正したものに，自由度調整済決定係数 （$Adjusted\ R^2$）がある．q を分割表のセル数，r と r_0 をそれぞれモデル X, X_0 の自由度とする．このとき，上の R^2 を

$$Adj.R^2 = 1 - \frac{q - r_0}{q - r}[1 - R^2]$$

と修正したものを自由度調整済決定係数という．モデルが大きい，言い換えると r が大きいほど，第２項の分母の $q - r$ により，$Adj.R^2$ の値は小さくなる．この値が大きいモデルほど良いと考える．また，もともとの R^2 の定義式から，

$$Adj.R^2 = 1 - \frac{G^2(X)/(q - r)}{G^2(X_0)/(q - r_0)}$$

とも変形できる．

赤池情報量規準 AIC

赤池情報量規準（Akaike's Information Criterion）は，元統計数理研究所所長の赤池弘次氏が1971 年に考案し 1973 年に発表したモデル評価規準であり，AIC とも呼ばれる．対数線形モデルに基づく分割表解析においては，

$$\mathrm{AIC} = G^2(X) - 2[q - r]$$

とし，これが小さいモデルが良いと判断する．モデルが当てはまりがよいなら $G^2(X)$ は小さくなるし，さらには，母数が少ないほど第 2 項の影響で AIC は小さくなる．道理にかなった規準であるといえる．

4.2 モデル選択過程

解析者にとって価値のある対数線形モデルというのは，当てはまりがよく，且つ，母数（効果項）の数もできる限り少なく抑えたモデルである．多元分割表で対数線形モデルを考える場合，初期モデルを仮定して，効果項を追加もしくは除去していき最良と考えられるモデルを導く手法がある．本節では，**ステップワイズ法**(stepwise procudures) を紹介する．この方法による効果の取捨は，検定統計量もしくはモデル選択規準の値をもとにして行う．ここでは，尤度比検定統計量 G^2 を利用する場合について記述する．

ステップワイズ法による効果項の取捨の方法には，追加型・除去型・混合型の 3 つの型がある．

追加型

効果項を水準値をもとに追加していく方法である．基本的な手順は，次のようになる．

1. 初期モデルを仮定する．
2. その時点のモデルに存在しない効果項のうち，最も有意である G^2 をもつ項をモデルに追加する．ただし，追加する効果は，その時点の効果より一段階大きな項に限定する．
3. 有意な G^2 をもつ効果項が存在しないとき，追加をやめる．

手順 2 に「一段階大きな項」とある．これは，例えば 3 元分割表モデル [1][2][3] が仮定されているときは，[12],[13],[23] のみを考え，[123] は考えないということである．4 元分割表モデル [1][23][4] が仮定されているときは，[12],[13],[14],[24],[34],[123],[234] のみを考えるということである．ここでの G^2 は,「項を追加したときのモデル」に対する現時点のモデルの尤度比検定統計量である．

除去型

効果項を水準値をもとに除去していく方法である．基本的な手順は，次のようになる．

1. 初期モデルを仮定する．
2. その時点のモデルの存在する極大項のうち，最小有意の G^2 をもつ項を除去する．
3. 極大項がすべて有意であるとき，除去をやめる．

ここでの G^2 は，現時点のモデルに対する「項を除去したときのモデル」の検定統計量である．

混合型

混合型は，1 回の取捨の度に追加型と除去型どちらが有効か判別し，有効な方を行う．追加・除去のどちらも行わないほうがモデルとしては良いとき，取捨をやめる．

初期モデル

ステップワイズ法を適用するためには，適切な初期モデルを設定しなければならない．最も単純な設定は，全 s 因子効果モデルに設定する方法である．この全 s 因子効果モデルとは，存在する因子により構成されうる全ての s 因子交互作用を備えたモデルという意味である．

例 4.2.1. 5 元分割表の対数線形モデルの場合，

- [1][2][3][4][5]
- [12][13][14][15][23][24][25][34][35][45]
- [123][124][125][134][135][145][234][235][245][345]
- [1234][1235][1345][2345]
- [12345]

の 5 タイプの全 s 因子効果モデルが構成される．

いくつかの全 s 因子効果モデルから，初期モデルをステップワイズの型に応じて選ぶ．通常は以下のように選ぶことが多い．

- 追加型ステップワイズでは，いくつかの全 s 因子効果モデルのうち，当てはまりが悪いモデルの中で最大のモデルを初期モデルとする．
- 除去型ステップワイズでは，いくつかの全 s 因子効果モデルのうち，当てはまりが良いモデルの中で最小のモデルを初期モデルとする．

- 混合型ステップワイズでは，上の 2 つのどちらかを初期モデルとする．

例 4.2.2. 5 元分割表の全 s 因子効果モデルのうち，検定統計量 G^2 により

当てはまり悪：
- [1][2][3][4][5]
- [12][13][14][15][23][24][25][34][35][45]
- [123][124][125][134][135][145][234][235][245][345]

当てはまり良：
- [1234][1235][1345][2345]
- [12345]

と判断されたとする．この結果から初期モデルをそれぞれ

追加型ステップワイズ： [123][124][125][134][135][145][234][235][245][345]

除去型ステップワイズ： [1234][1235][1345][2345]

混合型ステップワイズ： 上のどちらか

と設定し，効果項の取捨を行う．

4.3 データ解析例

次のデータは，虚血性疾患と肥満度との関係を調べるために実施された患者-対照研究の結果を，年齢と地域特性（寒冷地域，温暖地域）で分類しまとめたものである．こちらは参考文献 [10, p.149] より引用している．

年齢	地域特性	タイプ	肥満	正常	計
50〜59	寒冷	患者	56	22	78
		対照	12	78	90
	温暖	患者	66	19	85
		対照	15	84	99
60〜69	寒冷	患者	42	24	66
		対照	14	72	86
	温暖	患者	89	21	110
		対照	20	56	76
70〜79	寒冷	患者	50	46	96
		対照	31	135	166
	温暖	患者	52	33	85
		対照	11	37	48

この表は $3 \times 2 \times 2 \times 2$ の 4 元分割表である．各属性については，添字を

$$\text{年齢} : i, \text{地域特性} : j, \text{タイプ} : k, \text{体型} : l$$

と割り当て，属性番号は順に 1,2,3,4 とする．各セルは (i, j, k, l) と表される．調査の特性より，この表は $n_{hij\cdot}$ が固定である積多項（二項）表である．

このデータに対し，ステップワイズ法によるモデル選択を行う．まず，初期モデルの候補である全 s 因子モデルの当てはめを行う．

```
> rdata <- read.csv("research-disease.csv",header=T)
> rdata
   Age Temperature     Type   Body   N
1   A1        Cold  Patient    Fat  56
2   A1        Cold  Patient Normal  22
3   A1        Cold Contrast    Fat  12
4   A1        Cold Contrast Normal  78
5   A1        Warm  Patient    Fat  66
6   A1        Warm  Patient Normal  19
7   A1        Warm Contrast    Fat  15
8   A1        Warm Contrast Normal  84
9   A2        Cold  Patient    Fat  42
10  A2        Cold  Patient Normal  24
11  A2        Cold Contrast    Fat  14
12  A2        Cold Contrast Normal  72
13  A2        Warm  Patient    Fat  89
14  A2        Warm  Patient Normal  21
15  A2        Warm Contrast    Fat  20
16  A2        Warm Contrast Normal  56
17  A3        Cold  Patient    Fat  50
18  A3        Cold  Patient Normal  46
19  A3        Cold Contrast    Fat  31
20  A3        Cold Contrast Normal 135
21  A3        Warm  Patient    Fat  52
22  A3        Warm  Patient Normal  33
23  A3        Warm Contrast    Fat  11
24  A3        Warm Contrast Normal  37
>
> m1 <- loglm(N~Age+Temperature+Type+Body,data=rdata)
        #全１因子効果モデル
>
> m2 <- loglm(N~Age*Temperature+Age*Type+Age*Body+Temperature*Type
                +Temperature*Body+Type*Body,data=rdata)
```

```
        #全２因子効果モデル
>
> m3 <- loglm(N~Age*Temperature*Type+Age*Temperature*Body
                +Age*Type*Body+Temperature*Type*Body,data=rdata)
        #全３因子効果モデル
>
> m4 <- loglm(N~Age*Temperature*Type*Body,data=rdata)
        #全４因子効果モデル
>
>
> m1   #モデル１:[1][2][3][4]
Call:
loglm(formula = N ~ Age + Temperature + Type + Body, data = rdata)

Statistics:
                      X^2 df P(> X^2)
Likelihood Ratio 403.4988 18        0   #当てはまり悪
Pearson          427.6113 18        0
>
> m2   #モデル２:[12][13][14][23][24][34]
Call:
loglm(formula = N ~ Age * Temperature + Age * Type + Age * Body +
    Temperature * Type + Temperature * Body + Type * Body, data = rdata)

Statistics:
                      X^2 df     P(> X^2)
Likelihood Ratio 30.93248  9 0.0003040313   #当てはまり悪
Pearson          31.74922  9 0.0002200098
>
> m3   #モデル３:[123][124][134][234]
Call:
loglm(formula = N ~ Age * Temperature * Type + Age * Temperature *
    Body + Age * Type * Body + Temperature * Type * Body, data = rdata)

Statistics:
```

```
                        X^2 df  P(> X^2)
Likelihood Ratio 0.05256954  2 0.9740577   #当てはまり良
Pearson          0.05264788  2 0.9740195
>
> m4   #モデル４:[1234]（飽和モデル）
Call:
loglm(formula = N ~ Age * Temperature * Type * Body, data = rdata)

Statistics:
                 X^2 df P(> X^2)
Likelihood Ratio   0  0        1   #当てはまり良
Pearson            0  0        1
```

当てはまりの悪いモデルで，大きいのものはモデル２であるから，それを初期モデルにし，追加型ステップワイズを実行する．R の関数 step は AIC をもとに変数選択を行う関数である．候補効果項を追加（もしくは除去）したときの AIC を算出し，AIC が低い項をモデルに追加（もしくは除去）するという手順を繰り返し，これ以上 AIC が低い値にならなくなったとき選択をやめる．これらの一連の動作が自動的に行われる．

```
> #追加型ステップワイズ（初期モデル m2）
> step(m2,direction="forward",   #"forward"は追加型の指定
          scope=list(upper=~Age*Temperature*Type*Body  #追加候補の項
                           +Age*Temperature*Type+Age*Temperature*Body
                           +Age*Type*Body+Temperature*Type*Body))

Start:  AIC= 60.93   #初期モデル m2 の AIC
 N ~ Age * Temperature + Age * Type + Age * Body + Temperature *
    Type + Temperature * Body + Type * Body

                          Df    AIC   #"AIC"はその項を追加したときの値 AIC
+ Age:Temperature:Type     2 51.081     項 [123] を追加したとき AIC は 51.081,
+ Age:Type:Body            2 52.804     AIC は最小である.
+ Age:Temperature:Body     2 58.729   #左の"+"は追加の意味である.
<none>                       60.932   #"none"は「何もしなければ」の意味,
+ Temperature:Type:Body    1 61.857     つまり現時点のモデルのことである.
```

```
Step:  AIC= 51.08
 N ~ Age + Temperature + Type + Body + Age:Temperature + Age:Type +
    Age:Body + Temperature:Type + Temperature:Body + Type:Body +
    Age:Temperature:Type
#モデルに項 [123]（Age:Temperature:Type）が追加された.

                         Df    AIC
+ Age:Type:Body           2 40.654   #AIC 最小項:[134]
<none>                      51.081
+ Temperature:Type:Body   1 51.666
+ Age:Temperature:Body    2 51.871

Step:  AIC= 40.65
 N ~ Age + Temperature + Type + Body + Age:Temperature + Age:Type +
    Age:Body + Temperature:Type + Temperature:Body + Type:Body +
    Age:Temperature:Type + Age:Type:Body
#項 [134] を追加

                         Df    AIC
<none>                      40.654     #項を追加しないとき AIC が最小,
+ Temperature:Type:Body   1 42.367       ここでステップワイズを終了する.
+ Age:Temperature:Body    2 42.412
Call:
loglm(formula = N ~ Age + Temperature + Type + Body + Age:Temperature +
    Age:Type + Age:Body + Temperature:Type + Temperature:Body +
    Type:Body + Age:Temperature:Type + Age:Type:Body, data = rdata,
    evaluate = FALSE)
#選択された最終的なモデル（[24][123][134]）

Statistics:
                      X^2 df  P(> X^2)
Likelihood Ratio 2.654445  5 0.7530766   #最終的なモデルの検定統計量
Pearson          2.659215  5 0.7523489
```

以上より，追加型ステップワイズにより，

モデル : [24][123][134]

が良いモデルであると判断された.

次に，除去型ステップワイズを行う．初期モデルは，当てはまりの良いモデルで最小のモデル３にする．

```
> #除去型ステップワイズ（初期モデル m3）
> step(m3,direction="backward")   #"backward"で除去型の指定
Start:  AIC= 44.05
 N ~ Age * Temperature * Type + Age * Temperature * Body + Age *
    Type * Body + Temperature * Type * Body

                         Df    AIC
- Age:Temperature:Body    2 42.367   #[124] を除去したとき AIC 最小
- Temperature:Type:Body   1 42.412     "-"は除去の意味
<none>                      44.053
- Age:Temperature:Type    2 51.357
- Age:Type:Body           2 52.254

Step:  AIC= 42.37
 N ~ Age + Temperature + Type + Body + Age:Temperature + Age:Type +
    Temperature:Type + Age:Body + Temperature:Body + Type:Body +
    Age:Temperature:Type + Age:Type:Body + Temperature:Type:Body
#[124] を除去

                         Df    AIC
- Temperature:Type:Body   1 40.654   #AIC 最小項 [234]
<none>                      42.367
- Age:Type:Body           2 51.666
- Age:Temperature:Type    2 54.588

Step:  AIC= 40.65
 N ~ Age + Temperature + Type + Body + Age:Temperature + Age:Type +
    Temperature:Type + Age:Body + Temperature:Body + Type:Body +
    Age:Temperature:Type + Age:Type:Body
#[234] を除去
```

```
                        Df    AIC
<none>                     40.654   #現時点がAIC最小
- Temperature:Body       1 47.579
- Age:Type:Body          2 51.081
- Age:Temperature:Type   2 52.804
Call:
loglm(formula = N ~ Age + Temperature + Type + Body + Age:Temperature +
    Age:Type + Temperature:Type + Age:Body + Temperature:Body +
    Type:Body + Age:Temperature:Type + Age:Type:Body, data = rdata,
    evaluate = FALSE)
#選択された最終的なモデル

Statistics:
                      X^2 df  P(> X^2)
Likelihood Ratio 2.654445  5 0.7530766
Pearson          2.659215  5 0.7523489
```

以上より，除去型ステップワイズでも

モデル : [24][123][134]

が良いモデルであると判断された.

最後に混合型ステップワイズを行う．初期モデルをモデル２，モデル３に設定した両方の場合を確認する.

```
>#混合型ステップワイズ（m3を初期モデルとしたとき）
> step(m3,direction="both")   #"both"で混合型の指定
Start:  AIC= 44.05
 N ~ Age * Temperature * Type + Age * Temperature * Body + Age *
    Type * Body + Temperature * Type * Body

                         Df    AIC
- Age:Temperature:Body    2 42.367   #[124]の除去がAIC最小
- Temperature:Type:Body   1 42.412
<none>                      44.053
- Age:Temperature:Type    2 51.357
```

```
- Age:Type:Body          2 52.254

Step:  AIC= 42.37
 N ~ Age + Temperature + Type + Body + Age:Temperature + Age:Type +
    Temperature:Type + Age:Body + Temperature:Body + Type:Body +
    Age:Temperature:Type + Age:Type:Body + Temperature:Type:Body

                        Df    AIC
- Temperature:Type:Body  1 40.654   #[234] の除去が AIC 最小
<none>                     42.367
+ Age:Temperature:Body   2 44.053
- Age:Type:Body          2 51.666
- Age:Temperature:Type   2 54.588

Step:  AIC= 40.65
 N ~ Age + Temperature + Type + Body + Age:Temperature + Age:Type +
    Temperature:Type + Age:Body + Temperature:Body + Type:Body +
    Age:Temperature:Type + Age:Type:Body

                        Df    AIC
<none>                     40.654   #現段階が AIC 最小
+ Temperature:Type:Body  1 42.367
+ Age:Temperature:Body   2 42.412
- Temperature:Body       1 47.579
- Age:Type:Body          2 51.081
- Age:Temperature:Type   2 52.804
Call:
loglm(formula = N ~ Age + Temperature + Type + Body + Age:Temperature +
    Age:Type + Temperature:Type + Age:Body + Temperature:Body +
    Type:Body + Age:Temperature:Type + Age:Type:Body, data = rdata,
    evaluate = FALSE)

Statistics:
                      X^2 df  P(> X^2)
Likelihood Ratio 2.654445  5 0.7530766
```

```
Pearson          2.659215  5 0.7523489
>#モデル [24][123][134] が選択された
>
>
>#混合型ステップワイズ（m2 を初期モデルとしたとき）
> step(m2,direction="both",
       scope=list(upper=~Age*Temperature*Type*Body
                         +Age*Temperature*Type+Age*Temperature*Body
                         +Age*Type*Body+Temperature*Type*Body))
Start:  AIC= 60.93
 N ~ Age * Temperature + Age * Type + Age * Body + Temperature *
    Type + Temperature * Body + Type * Body

                         Df    AIC
+ Age:Temperature:Type    2  51.08   #[123] の追加が AIC 最小
+ Age:Type:Body           2  52.80
- Age:Type                2  58.39
+ Age:Temperature:Body    2  58.73
<none>                       60.93
+ Temperature:Type:Body   1  61.86
- Age:Body                2  63.01
- Temperature:Body        1  65.46
- Temperature:Type        1  66.26
- Age:Temperature         2  94.15
- Type:Body               1 333.71

Step:  AIC= 51.08
 N ~ Age + Temperature + Type + Body + Age:Temperature + Age:Type +
    Age:Body + Temperature:Type + Temperature:Body + Type:Body +
    Age:Temperature:Type

                         Df    AIC
+ Age:Type:Body           2  40.65   #[134] の追加が AIC 最小
<none>                       51.08
+ Temperature:Type:Body   1  51.67
```

```
+ Age:Temperature:Body   2  51.87
- Age:Body               2  53.33
- Temperature:Body       1  55.77
- Age:Temperature:Type   2  60.93
- Type:Body              1 324.03

Step:  AIC= 40.65
 N ~ Age + Temperature + Type + Body + Age:Temperature + Age:Type +
    Age:Body + Temperature:Type + Temperature:Body + Type:Body +
    Age:Temperature:Type + Age:Type:Body

                        Df    AIC
<none>                     40.654   #現段階が AIC 最小
+ Temperature:Type:Body  1 42.367
+ Age:Temperature:Body   2 42.412
- Temperature:Body       1 47.579
- Age:Type:Body          2 51.081
- Age:Temperature:Type   2 52.804
Call:
loglm(formula = N ~ Age + Temperature + Type + Body + Age:Temperature +
    Age:Type + Age:Body + Temperature:Type + Temperature:Body +
    Type:Body + Age:Temperature:Type + Age:Type:Body, data = rdata,
    evaluate = FALSE)

Statistics:
                      X^2 df  P(> X^2)
Likelihood Ratio 2.654445  5 0.7530766
Pearson          2.659215  5 0.7523489
>#モデル [24][123][134] が選択された
```

以上より，混合型でも

モデル : [24][123][134]

が良いと判断された.

第5章　グラフィカルモデリング

グラフィカルモデリングとは，多変量データの変数関連構造を推測し，グラフでその構造を表現する手法であり，比較的最近開発された多変量解析法の１つである．本章では，多元分割表の対数線形モデルのグラフ表現と，グラフをもとにしたモデル選択について簡潔に記述する．なお，本章では対象の表を多項サンプリング表に限定しておく．

グラフ理論用語

対数線形モデルのグラフ表現を行う上で必要なグラフ理論の用語を簡単に紹介しておく．

頂点からなる集合 V と，頂点を結ぶ**辺**からなる集合 E からなる構造 $G=(V,E)$ を**グラフ**という．

$(v_1,v_2)\in E$ のとき，すなわち，頂点 v_1,v_2 を結ぶ辺が存在するとき，v_1と v_2 は**隣接している**という．また，すべての頂点の対が辺で結ばれているとき，グラフは**完全**であるという．

グラフの辺に方向を考えないとき，そのグラフのことを**無向グラフ**という．

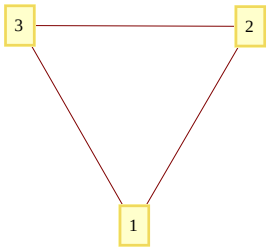
頂点集合 V の部分集合 V_1 に対して，V_1 に属す頂点と両端が V_1 に属す辺からなるグラフを，V_1 からなる G の**部分グラフ**といい，これを G_{V_1} と表す．$V_2\subseteq V$ において，V_2 が生成する部分グラフが極大且つ完全であるとき，すなわち G_{V_2} が完全で，$V_1\supset V_2(V_1\neq V_2)$ である任意の V_1 の G_{V_1} が完全でないとき，V_2 を**クリーク**という．

異なる頂点の列 $v_0,v_1,\ldots,v_n$ は，任意の $j=1,\ldots,n$ について $(v_{j-1},v_j)\in E$ のとき，長さ n の**道**という．頂点 v_i,v_j を含む道があるとき，v_i,v_j は**連結している**という．長さ n の道 $v_0,v_1,\ldots,v_n$ で，$v_0=v_n$ を許したものを，長さ n の**閉路**という．長さ n の閉路で，連続していない頂点を結ぶ辺を**弦**という．長さが 4 以上の任意の閉路に弦が存在するとき，そのグラフは**三角化している**という．

5.1　対数線形モデルのグラフ表現

対数線形モデルのグラフ表現の基本的な方法は，モデルのもとである分割表の属性を頂点，モデルの 2 因子交互作用を辺に対応させてグラフを書くというものである．

例 5.1.1. 3 元分割表の対数線形モデル $\log(m_{ijk})=u+u_{1(i)}+u_{2(j)}+u_{3(k)}+u_{12(ij)}+u_{13(ik)}+u_{23(jk)}$ は生成集合表現で [12][13][23] と書ける．これをグラフで表すと，



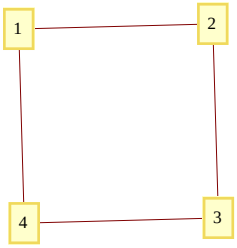
となる．3 つの因子が頂点，3 つの交互作用が辺で表わされる．

以下，効果項を，u_{12} のように u と属性番号で表すことにする．モデルのグラフ表現法を一般化し，無向独立グラフを定義する．

定義 5.1.1. q 個の確率変数 $X_i, i = 1, \ldots, q$ があるとする．ある確率変数の対 (X_j, X_k) が，他の $q-2$ 個の変数をどのように与えても条件付独立でないとき，その X_j, X_k を表わす頂点間を辺で結ぶ．このようにして構成される無向グラフを**無向独立グラフ**という．

この定義はつまり，属性 X_i, X_j の交互作用 u_{ij} が存在するとき，その点同士を結ぶ，こうして構成されたグラフが無向独立グラフであるということである．

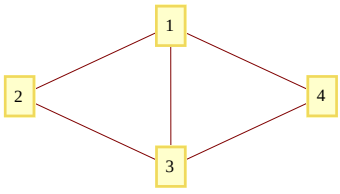
例 5.1.2. モデル [12][14][23][34] は，



という無向独立グラフで表せる．

別の例を与える．

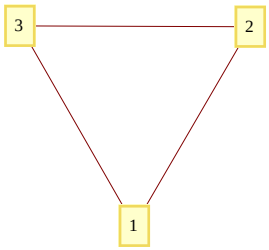
例 5.1.3. モデル [12][13][14][23][34] とモデル [123][134] は，両方とも $u_{12}, u_{13}, u_{14}, u_{23}, u_{34}$ という交互作用を含んでいるため，同じ無向独立グラフで表される．



この例のように，無向独立グラフと対数線形モデルは 1 対 1 に対応するわけではない．この問題を解消するために，次のようなモデルに対象を限定する．

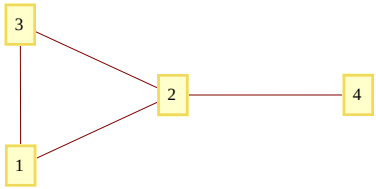
定義 5.1.2. 与えられた無向グラフに対して，グラフのクリークの集合が生成集合となる対数線形モデルを**グラフィカル対数線形モデル**という．

例 5.1.4. 3 元分割表の対数線形モデルで，[123] はグラフィカル対数線形モデルであるが，[12][13][23] はそうではない．どちらのモデルも u_{12}, u_{13}, u_{23} が存在するが，グラフィカルなモデルとなるためにはそれらの項が生成する u_{123} が存在しなければならない．[123] のグラフは，



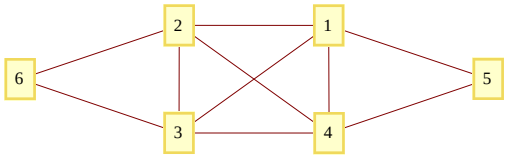
である．クリークは $\{1,2,3\}$ であり，モデル [123] の生成集合と一致する．

例 5.1.5. 4 元分割表の対数線形モデルで，[123][24] はグラフィカル対数線形モデルであるが，[12][13][23][24] はそうではない．どちらのモデルも無向独立グラフは，



である．クリークは $\{1,2,3\}, \{2,4\}$ の 2 種類が存在し，これと一致するのは [123][24] の生成集合である．後者のモデルは u_{12}, u_{13}, u_{23} が存在するから，グラフィカルとなるためには u_{123} も存在する必要がある．

例 5.1.6. 6 元分割表の [123][124][134][234][145][236] というモデルのグラフは，

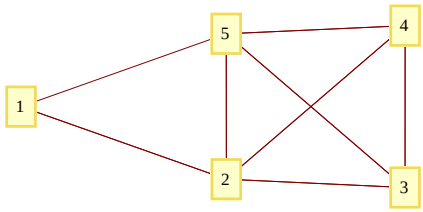


と書けるが，このグラフのクリークは $\{1,2,3,4\}, \{1,4,5\}, \{2,3,6\}$ であるため，モデルはグラフィカルではない．グラフィカルとなるためには u_{1234} が必要となる．

グラフィカルモデルになるためには，ある高次元の交互作用を生成する一次元小さい交互作用が全て存在するとき，その高次元の交互作用がモデルに含まれなければならない．この高次元の交互作用が存在しなければならないという条件が階層モデルとの違いである．グラフィカル対数線形モデルは，階層モデルの部分クラスとなる．

グラフィカル対数線形モデルに対象を絞ることで，モデルをグラフで表現するだけでなく，逆にグラフから対数線形モデルを読み取ることも可能である．

例 5.1.7. 次のグラフを与える．



この場合，クリークは $\{1,2,5\},\{2,3,4,5\}$ であるので，対応するグラフィカル対数線形モデルは $[125][2345]$ である．

モデルをグラフ表現する最大の利点というのは，条件付独立関係を読み取りやすい点である．

定理 5.1.1. [6, p187]
集合 V_1,V_2,V_3 が，グラフィカル対数線形モデルの頂点集合の互いに排反の部分集合のとき，「V_3 が与えられたもとで V_1 と V_2 は独立」と「V_1,V_2 間の任意の道に少なくとも 1 つの V_3 の頂点が含まれる」は同値である．

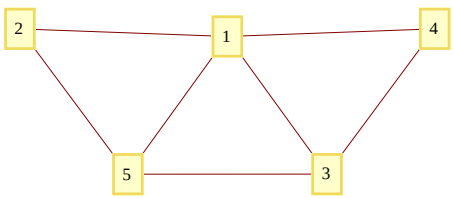
例 5.1.8. 次のグラフを与える．

$V_1=\{1,2\},V_2=\{5,6\},V_3=\{3,4\}$ とすると，V_1 と V_2 を繋ぐ道には，V_3 の点が存在する．よって，$(X_1,X_2)\perp\!\!\!\perp(X_5,X_6)\mid(X_3,X_4)$ である．
V_i のとり方は何通りもあり，例えば $V_1=\{1\},V_2=\{5,6\},V_3=\{2,3,4\}$ のようにもとれるが，前と同様に条件付独立関係が成り立つ．
元のモデルは，$[123][456][34]$ である．

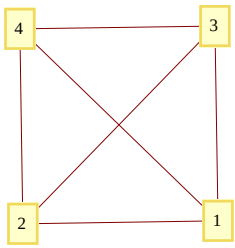
さらに，グラフィカル対数線形モデルの部分クラスを与える．

定義 5.1.3. グラフィカル対数線形モデルのグラフが三角化しているとき，**分解可能モデル**という．

例 5.1.9. モデル $[125][134][135]$ はグラフィカルモデルである．そのグラフは，

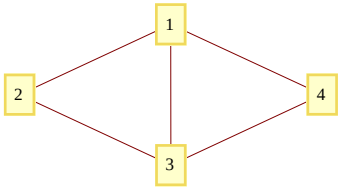


である．これには，長さ 4 以上のどの閉路にも近道である弦が存在する．よってこのモデルは分解可能モデルである．
グラフィカルではないが三角化しているモデルは存在する．モデル [12][13][14][23][24][34] は，グラフィカルではない．そのグラフは，



であるが，長さ 4 以上の閉路には弦が存在している．よって三角化している．ただし，元々がグラフィカル対数線形モデルではないので，分解可能モデルではない．

例 5.1.10. 次のグラフ



は分解可能である．モデルは [123][134] であり，セル確率は，

$$p_{ijkl} = \frac{p_{ijk\cdot}p_{i\cdot kl}}{p_{i\cdot k\cdot}}$$

と書ける．これより，多項表のとき

$$\begin{aligned}\hat{m}_{ijkl} &= n_{\cdots\cdot}\hat{p}_{ijkl} \\ &= n_{\cdots\cdot}\left(\frac{n_{ijk\cdot}}{n_{\cdots\cdot}}\cdot\frac{n_{i\cdot kl}}{n_{\cdots\cdot}}\Big/\frac{n_{i\cdot k\cdot}}{n_{\cdots\cdot}}\right) \\ &= \frac{n_{ijk\cdot}n_{i\cdot kl}}{n_{i\cdot k\cdot}}\end{aligned}$$

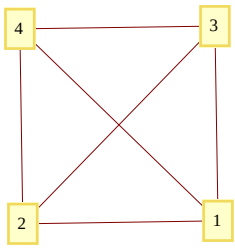
と書ける．
p_{ijkl} と $\hat{m}_{ijkl}$ はそれぞれ，分母は 2 つのクリークの共通部分である $\{1,3\}$ に依存，分子にクリーク $\{1,2,3\},\{1,3,4\}$ に依存する項からなる．

このように，分解可能モデルは，属性関連が解釈しやすいように分解されるという利点があるため，グラフィカルモデリングにおいて分解可能モデルに限定する場合は多い．

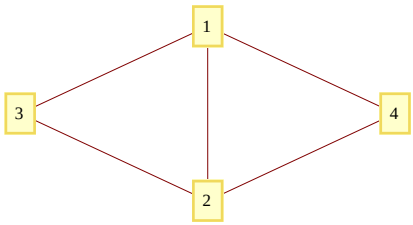
5.2　分解可能モデル選択過程

対象を分解可能モデルに限定する．グラフを用いたモデル選択としては，完全グラフ，つまり飽和モデルのグラフから不必要な辺を１辺ずつ除去していく方針が考えられる．ただし，除去の対象となる辺は，グラフにおいて２つ以上のクリークに属す辺以外である必要がある．

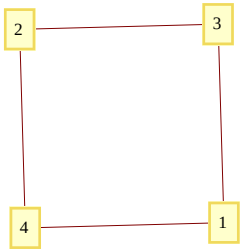
例 5.2.1. 飽和モデル [1234] のグラフは，



であり，ここから辺 (3,4) が除去されたとする．そうすると，



である．このモデルは [123][124] である．辺 (1,2) は２つのクリークに属している．さらに，辺の除去を考えるとき，もし辺 12 を除去すると



となり，分解可能でなくなってしまう．

辺の除去後も分解可能であるためには，２つ以上のクリークに属す辺を除去を対象にしない必要がある．

辺の除去の指針であるが，除去対象の任意の辺について,「元のモデル」に対する「辺を除去したときのモデル」の検定を行い，P 値が最大になった辺を除去する手順が有効である．
飽和モデルに対する「辺を除去したときのモデル」の検定を行う方法もある．

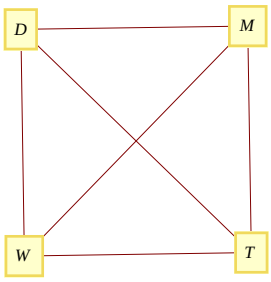
5.3　データ解析例

次のデータは，ある研究において，マウスに 2 種類の薬品を投与したときの筋張力の変化を調べた結果で，元々のマウスの筋肉量とタイプを含む 4 属性で分類したものである．こちらは参考文献 [6, p.109] より引用している．

			薬品	
筋張力の変化	筋肉量	筋肉	薬品 1	薬品 2
High	High	Type 1	3	21
		Type 2	23	11
	Low	Type 1	22	32
		Type 2	4	12
Low	High	Type 1	3	10
		Type 2	41	21
	Low	Type 1	45	23
		Type 2	6	22

各属性の表記は，筋張力の変化：T，筋肉量：W，筋肉：M，薬品：D と与える．

飽和モデル [TWMD] のグラフは次のようになる．



ここから辺の除去を行う．この段階では，全ての辺が除去対象である．R にて，各辺（2 因子交互作用）を除去したときのモデルの G^2，P 値を確認する．

```
> rdata <- read.csv("R-md-data.csv",header=T)
> rdata
      T    W  M  D  N
1  HIGH high t1 d1  3
2  HIGH high t1 d2 21
3  HIGH high t2 d1 23
4  HIGH high t2 d2 11
5  HIGH  low t1 d1 22
6  HIGH  low t1 d2 32
```

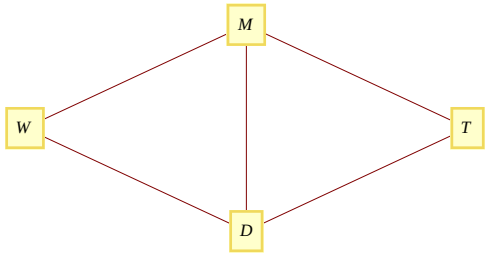
```
7  HIGH  low t2 d1  4
8  HIGH  low t2 d2 12
9   LOW high t1 d1  3
10  LOW high t1 d2 10
11  LOW high t2 d1 41
12  LOW high t2 d2 21
13  LOW  low t1 d1 45
14  LOW  low t1 d2 23
15  LOW  low t2 d1  6
16  LOW  low t2 d2 22
>
>
> m1 <- loglm(N~T*W*M+T*W*D,data=rdata) #辺 MD 除去
> m2 <- loglm(N~T*W*M+T*M*D,data=rdata) #辺 WD 除去
> m3 <- loglm(N~T*W*M+W*M*D,data=rdata) #辺 TD 除去
> m4 <- loglm(N~T*W*D+T*M*D,data=rdata) #辺 WM 除去
> m5 <- loglm(N~T*W*D+W*M*D,data=rdata) #辺 TM 除去
> m6 <- loglm(N~T*M*D+W*M*D,data=rdata) #辺 TW 除去
>
> m1
Call:
loglm(formula = N ~ T * W * M + T * W * D, data = rdata)

Statistics:
                     X^2 df     P(> X^2)
Likelihood Ratio 45.13186  4 3.732715e-09
Pearson          42.77174  4 1.154009e-08
>
> # m2 から m6 も同様に確認する
```

次の表は，R で確認した辺を除去した 6 つのモデルの G^2，P 値をまとめたものである．

R	除去した辺	モデル	自由度	G^2	P 値
m1	MD	[TWM][TWD]	4	45.132	0
m2	WD	[TWM][TMD]	4	39.503	0
m3	TD	[TWM][WMD]	4	8.693	0.0692
m4	WM	[TWD][TMD]	4	107.059	0
m5	TM	[TWD][WMD]	4	12.251	0.0168
m6	TW	[TMD][WMD]	4	1.529	0.8215

この結果より，P 値が最大である辺 TW を除去することにする．飽和モデルのグラフから辺 TW を除去したものと同値なグラフを与える．



現時点のモデルは [TMD][WMD] である．

さらに辺の除去を考える．候補となる辺は，２つのクリークに共通する辺 MD 以外の４つの辺，WD,WM,TD,TM である．先と同様に R を実行する．

```
> m6.1 <- loglm(N~T*M*D+W*M,data=rdata) #辺 WD 除去
> m6.2 <- loglm(N~T*M*D+W*D,data=rdata) #辺 WM 除去
> m6.3 <- loglm(N~T*M+W*M*D,data=rdata) #辺 TD 除去
> m6.4 <- loglm(N~T*D+W*M*D,data=rdata) #辺 TM 除去
>
> m6.1
Call:
loglm(formula = N ~ T * M * D + W * M, data = rdata)

Statistics:
                      X^2 df     P(> X^2)
Likelihood Ratio 44.38897  6 6.188767e-08
Pearson          41.83474  6 1.982217e-07
>
> # 同様に m6.2，m6.3，m6.4 も確認する
```

次の表では，各モデルの G^2 と，現時点のモデル [TMD][WMD] の G^2 との差と，その P 値をまとめてある．

R	除去した辺	モデル	自由度	G^2	P 値	[TMD][WMD] との差 自由度	G^2	P 値
m6	-	[TMD][WMD]	4	1.529	0.8215	-	-	-
m6.1	WD	[TMD][WM]	6	44.389	0	2	42.86	0
m6.2	WM	[TMD][WD]	6	107.269	0	2	105.74	0
m6.3	TD	[WMD][TM]	6	13.579	0.0347	2	12.05	0.00242
m6.4	TM	[WMD][TD]	6	12.461	0.0524	2	10.932	0.00423

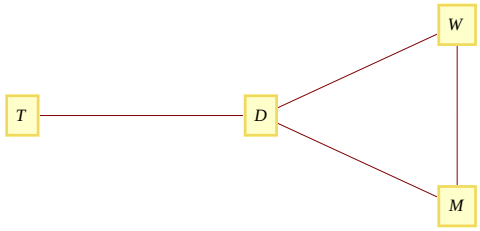
G^2 の差の P 値も R で算出している．m6.3 の場合，G^2 の差の値 12.05 を用いて，

```
> 1-pchisq(12.05,2)
[1] 0.002417552
```

としている．”pchisq(12.05,2)”は自由度 2 のカイ二乗分布の下側確率 $\mathrm{P}(\chi^2 \leq 12.05)$ という意味の R の関数である．この値こそ P 値である．

差の検定統計量を重視，つまり，現時点のモデルをより縮小できるかどうかの検定を重視するならば，差の P 値を見る限り辺の除去はこれ以上行うべきではないと判断する．

飽和モデルに対する検定の G^2 の P 値の中では，辺 TM を除去した [WMD][TD] が 0.05 を超えて最大であるから，試しにこのモデルを仮定してみる．現時点のグラフは次のようになる．



ここからさらに辺を除去したい．任意の辺について除去される可能性がある．先と同様に R を実行する．

```
> m6.4.1 <- loglm(N~T*D+W*M+W*D,data=rdata) #辺 MD 除去
> m6.4.2 <- loglm(N~T*D+W*M+M*D,data=rdata) #辺 WD 除去
> m6.4.3 <- loglm(N~T*D+W*D+M*D,data=rdata) #辺 WM 除去
> m6.4.4 <- loglm(N~T+W*M*D,data=rdata)     #辺 TD 除去
>
> m6.4.1
Call:
loglm(formula = N ~ T * D + W * M + W * D, data = rdata)

Statistics:
                      X^2 df     P(> X^2)
```

```
Likelihood Ratio 55.60526  8 3.366830e-09
Pearson          55.77386  8 3.121818e-09
>
> # 同様に他のモデルも確認
```

先ほどと同様の表で結果をまとめる.

						[WMD][TD] との差		
R	除去した辺	モデル	自由度	G^2	P 値	自由度	G^2	P 値
m6.4	-	[WMD][TD]	6	12.461	0.0524	-	-	-
m6.4.1	MD	[TD][WM][WD]	8	55.605	0	2	43.144	0
m6.4.2	WD	[TD][WM][MD]	8	55.321	0	2	42.86	0
m6.4.3	WM	[TD][WD][MD]	8	118.201	0	2	105.74	0
m6.4.4	TD	[WMD][T]	7	19.019	0.0081	1	6.558	0.01044

除去する辺は，強いて言うならば辺 TD であるが，２種類の P 値はどちらも 0.001 ほどである．
これを除去するかどうかは解析者の判断によるところが大きい．

付 録A　比例反復法プログラム

以下は，3元分割表モデル $M^{(8)}$ のもとでの期待度数の MLE を，比例反復法により算出するための C 言語を用いたプログラムの自作のソースコードである．これは第 3.3.2 節の手順に即して作成した．なお，$M^{(8)}$ と同値なモデル $M_*^{(8)}$ は，

$$\log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)}$$

であり，周辺制約は，

$$\hat{m}_{ij\cdot} = n_{ij\cdot}, \hat{m}_{i\cdot k} = n_{i\cdot k}, \hat{m}_{\cdot jk} = n_{\cdot jk}$$

であった．

```
------------------------------------------------------
#include<stdio.h>
#include<math.h>

main(){
  int i,j,k,p,q,r,t;
  int a[10][10];  /*周辺和 n_{ij.} */
  int b[10][10];  /*周辺和 n_{i.k} */
  int c[10][10];  /*周辺和 n_{.jk} */
  double x[10][10];
  int data[10][10][10];
  double m[10][10][10]; /

  printf("サイズの入力\n");
  printf("行--->");
  scanf("%d",&p);   /* 行数入力 */
  printf("列--->");
  scanf("%d",&q);   /* 列数入力 */
  printf("層--->");
  scanf("%d",&r);   /* 層数入力 */
```

```
printf("データの入力\n");

for(i = 0; i < p; i++){
  for(j = 0; j < q; j++){
    for(k = 0; k < r; k++){
      printf("data[%d][%d][%d]--->", i+1,j+1,k+1);
      scanf("%d",&data[i][j][k]); /* 度数入力 */
    }
  }
}
for(k = 0; k < r; k++){
  for(i = 0; i < p; i++){
    for(j = 0; j < q; j++){
      printf("data[%d][%d][%d]=%d \t", i+1,j+1,k+1,data[i][j][k]);
      /*元のデータ*/
    }
  printf("\n");
  }
printf("\n");
}

getchar();

/*周辺和の算出*/
for(i=0;i<p;i++){
  for(j=0;j<q;j++){
    a[i][j]=0;
    for(k=0;k<r;k++){
    a[i][j] += data[i][j][k]; /* 周辺和 n_{ij.} */
    }
  }
}
for(i=0;i<p;i++){
  for(k=0;k<r;k++){
    b[i][k]=0;
```

```
    for(j=0;j<q;j++){
    b[i][k] += data[i][j][k]; /* 周辺和 n_{i.k} */
    }
  }
}
for(j=0;j<q;j++){
  for(k=0;k<r;k++){
    c[j][k]=0;
    for(i=0;i<p;i++){
    c[j][k] += data[i][j][k]; /* 周辺和 n_{.jk} */
    }
  }
}

/*周辺和の表示*/
printf("\n n_{ij.}\n");
for(i = 0; i < p; i++){
  for(j = 0; j < q; j++){
    printf("a[%d][%d]=%d \t", i+1,j+1,a[i][j]);
  }
  printf("\n");
}
printf("\n n_{i.k}\n");
for(i = 0; i < 2; i++){
  for(k = 0; k < r; k++){
    printf("b[%d][%d]=%d \t", i+1,k+1,b[i][k]);
  }
  printf("\n");
}
printf("\n n_{.jk}\n");
for(j = 0; j < q; j++){
  for(k = 0; k < r; k++){
    printf("c[%d][%d]=%d \t", j+1,k+1,c[j][k]);
  }
  printf("\n");
```

```
}
printf("\n");

for(i = 0; i < p; i++){
  for(j = 0; j < q; j++){
    for(k = 0; k < r; k++){
      m[i][j][k]=1; /*初期値の設定*/
    }
  }
}

/*反復法の開始*/
for(t=0;t<10000;t++){      /*反復回数*/

  /*{ij.}による修正（3t+1 回目）*/
  for(i=0;i<p;i++){
    for(j=0;j<q;j++){
      x[i][j]=0;
    }
  }
  for(i=0;i<p;i++){
    for(j=0;j<q;j++){
      for(k=0;k<r;k++){
      x[i][j]+=m[i][j][k];
      }
    }
  }
  for(k = 0; k < r; k++){
    for(j = 0; j < q; j++){
      for(i = 0; i < p; i++){
        m[i][j][k]=a[i][j]*m[i][j][k]/x[i][j];
      }
    }
  }
  /*{i.k}による修正（3t+2 回目）*/
```

```
for(i=0;i<p;i++){
  for(k=0;k<r;k++){
    x[i][k]=0;
  }
}
for(i=0;i<p;i++){
  for(k=0;k<r;k++){
    for(j=0;j<q;j++){
    x[i][k]+=m[i][j][k];
    }
  }
}
for(k = 0; k < r; k++){
  for(j = 0; j < q; j++){
    for(i = 0; i < p; i++){
      m[i][j][k]=b[i][k]*m[i][j][k]/x[i][k];
    }
  }
}
/*{.jk}による修正（3(t+1)回目）*/
for(j=0;j<q;j++){
  for(k=0;k<r;k++){
    x[j][k]=0;
  }
}
for(j=0;j<q;j++){
  for(k=0;k<r;k++){
    for(i=0;i<p;i++){
    x[j][k]+=m[i][j][k];
    }
  }
}
for(k = 0; k < r; k++){
  for(j = 0; j < q; j++){
    for(i = 0; i < p; i++){
```

```
          m[i][j][k]=c[j][k]*m[i][j][k]/x[j][k];
        }
      }
    }
  }

  for(k = 0; k < r; k++){
    for(i = 0; i < p; i++){
      for(j = 0; j < q; j++){
        printf("m[%d][%d][%d]=%f \t", i+1,j+1,k+1,m[i][j][k]);
        /*最尤推定値 */
      }
    printf("\n");
    }
  printf("\n");
  }
  getchar();
  return 0;
}
```

実行例

次の表データに対して上のプログラムを実行してみる．こちらの表は参考文献 [6, p.73] より引用している．

タイプ	コレステロール値	最低血圧 (k)	
i	j	Normal	High
A	Nomal	716	79
	High	207	25
B	Normal	819	67
	High	186	22

サイズの入力

行--->2

列--->2

層--->2

データの入力

```
data[1][1][1]--->716
data[1][1][2]--->79
data[1][2][1]--->207
data[1][2][2]--->25
data[2][1][1]--->819
data[2][1][2]--->67
data[2][2][1]--->186
data[2][2][2]--->22

data[1][1][1]=716       data[1][2][1]=207
data[2][1][1]=819       data[2][2][1]=186

data[1][1][2]=79        data[1][2][2]=25
data[2][1][2]=67        data[2][2][2]=22


 n_{ij.}
a[1][1]=795     a[1][2]=232
a[2][1]=886     a[2][2]=208

 n_{i.k}
b[1][1]=923     b[1][2]=104
b[2][1]=1005    b[2][2]=89

 n_{.jk}
c[1][1]=1535    c[1][2]=146
c[2][1]=393     c[2][2]=47

m[1][1][1]=718.198774   m[1][2][1]=204.801226
m[2][1][1]=816.801226   m[2][2][1]=188.198774

m[1][1][2]=76.801226    m[1][2][2]=27.198774
m[2][1][2]=69.198774    m[2][2][2]=19.801226
-------------------------------------------------------
```

比較しやすいように，計算結果をはじめの表の形式でまとめた.

$\hat{m}_{ijk}$		k	
i	j	1	2
1	1	718.198774	76.801226
	2	204.801226	27.198774
2	1	816.801226	69.198774
	2	188.198774	19.801226

これらの数値が正しいかどうか確認するために，R でも算出してみる．R では，$M^{(8)}$ と同値な $M_*^{(8)}$ の当てはめを行っている．

```
> rdata <- read.csv("C-P-data.csv",header=T)
> rdata
  Type Cholesterol B.Pressure   N
1    A           N         PN 716
2    A           N         PH  79
3    A           H         PN 207
4    A           H         PH  25
5    B           N         PN 819
6    B           N         PH  67
7    B           H         PN 186
8    B           H         PH  22
> rdata.m <- loglm(N~Type*Cholesterol+Type*B.Pressure   #モデル [12][13][23]
                     +Cholesterol*B.Pressure,data=rdata)
> rdata.m
Call:
loglm(formula = N ~ Type * Cholesterol + Type * B.Pressure +
    Cholesterol * B.Pressure, data = rdata)

Statistics:
                      X^2 df  P(> X^2)
Likelihood Ratio 0.6132732  1 0.4335580
Pearson          0.6166685  1 0.4322880
> fitted(rdata.m)
Re-fitting to get fitted values
, , B.Pressure = PH   # k:High のセル

    Cholesterol
Type        H        N
```

```
   A 27.19877 76.80122
   B 19.80123 69.19878

, , B.Pressure = PN   # k:Normal のセル

    Cholesterol
Type        H        N
   A 204.8012 718.1988
   B 188.1988 816.8012
```

こちらも元の形式でまとめた.

$\hat{m}_{ijk}$		k	
i	j	1	2
1	1	718.1988	76.80122
	2	204.8012	27.19877
2	1	816.8012	69.19878
	2	188.1988	19.80123

結果を見比べるとほぼ一致しており，プログラムは $\hat{m}_{ijk}^{(8)}$ を算出するように正しく動いたようである.

他の３元分割表モデルのモデルや，４元以上の分割表についても，同様に作成できる.

参考文献

[1] 宮川 雅巳, 統計技法, 共立出版, 1998

[2] 宮川 雅巳, グラフィカルモデリング, 朝倉書店, 1997

[3] JST CREST 日比チーム, グレブナー道場, 共立出版, 2011

[4] 竹村 彰通, 現代数理統計学, 創文社, 1991

[5] 廣津千尋, 離散データ解析, 教育出版, 1982

[6] Ronald Christensen, Log-Linear Models and Logistic Regression, Springer, 1990

[7] JIN'S PAGE, http://mjin.doshisha.ac.jp/R/

[8] CHILD RESEARCH NET, http://www.blog.crn.or.jp

[9] 内閣府ホームページ, http://www.cao.go.jp/index.html

[10] 出村慎一, 健康・スポーツ科学のための統計学, 大修館書店, 1996