数理解析研究所講究録
第1758巻 2011年 184-194

On estimating multivariate discrete probabilities by pooled incomplete samples and related topics

関東学院大学　経済学部　布能 英一郎

1. Introduction

Proposition 1.1 k, m は $m < k$ なる自然数。確率変数 $\mathbf{X}$, $\mathbf{Y}$ は互いに独立で

$$\mathbf{X} = (X_1, \cdots, X_m, \cdots, X_k) \sim \text{Multinomial}(N_1; p_0, \cdots, p_m, \cdots, p_k),$$
$$\mathbf{Y} = (Y_1, \cdots, Y_m) \sim \text{Multinomial}(N_2; \frac{p_0}{\sum_{j=0}^{m} p_j}, \frac{p_1}{\sum_{j=0}^{m} p_j}, \cdots, \frac{p_m}{\sum_{j=0}^{m} p_j})$$

とする。この状況下で、Asano(1965) は p_i の MLE $\hat{p}_i$ が

$$\frac{x_i + y_i}{N_1\Big(1 + \frac{N_2}{\sum_{j=0}^{m} x_j}\Big)}, \quad \text{if } i \leq m, \qquad \frac{x_i}{N_1}, \quad \text{if } i > m \tag{1}$$

であることを示し、更に、$\hat{p}_i$ に関する諸性質を研究した。これと同様な現象は、他の分布の下でもいくつか生じている。

Proposition 1.2 確率変数 $\mathbf{X}$, $\mathbf{Y}$ は互いに独立で

$$\mathbf{X} = (X_1, \cdots, X_m, \cdots, X_k) \sim \text{NegativeMultinomial}(r_1; p_1, \cdots, p_m, \cdots, p_k),$$
$$\mathbf{Y} = (Y_1, \cdots, Y_m) \sim \text{NegativeMultinomial}(r_2; \frac{(1-p_0)p_1}{\sum_{j=1}^{m} p_j}, \cdots, \frac{(1-p_0)p_m}{\sum_{j=1}^{m} p_j})$$

ならば

$$\hat{p}_i = \begin{cases} \dfrac{T_x + T_y}{T_x + T_y + r_1 + r_2} \dfrac{x_i + y_i}{T_x\Big(1 + \dfrac{T_y}{\sum_{j=1}^{m} x_j}\Big)}, & 1 \leq i \leq m, \\ \dfrac{T_x + T_y}{T_x + T_y + r_1 + r_2} \dfrac{x_i}{T_x}, & i > m. \end{cases} \tag{2}$$

但し、$T_x = \sum_{j=1}^{k} x_j$, $T_y = \sum_{j=1}^{m} y_j$.

Proposition 1.3 $X_1, \cdots, X_m, \cdots, X_k, Y_1, \cdots, Y_m$ はすべて独立で,

$$X_i \sim Poisson(\lambda_i), \quad Y_i \sim Poisson((\lambda_1 + \cdots + \lambda_m + \cdots + \lambda_k)\frac{\lambda_i}{\lambda_1 + \cdots + \lambda_m})$$

と仮定。このとき、λ_i の MLE $\hat{\lambda}_i$ は

$$\hat{\lambda}_i = \begin{cases} \dfrac{T_x+T_y}{2}\dfrac{x_i+y_i}{T_x\Big(1+\dfrac{T_y}{\sum_{j=1}^m x_j}\Big)} & \text{if } i \le m, \\ \dfrac{T_x+T_y}{2}\dfrac{x_i}{T_x} & \text{if } i > m. \end{cases} \tag{3}$$

Proposition 1.2, 1.3 にて、MLE が Asano の形と同じようになることを、次表で見るとわかりやすい。

	$i \le m$ に対して	$i \ge m+1$ に対して
Proposition1.1 (多項分布)(Asano)	$\dfrac{x_i+y_i}{N_1\Big(1+\dfrac{N_2}{\sum_{j=0}^m x_j}\Big)}$	$\dfrac{x_i}{N_1}$
Proposition1.2 (負の多項分布)	$\dfrac{T_x+T_y}{T_x+T_y+r_1+r_2}\dfrac{x_i+y_i}{T_x\Big(1+\dfrac{T_y}{\sum_{j=1}^m x_j}\Big)}$	$\dfrac{T_x+T_y}{T_x+T_y+r_1+r_2}\dfrac{x_i}{T_x}$
Proposition1.3 (ポアソン分布)	$\dfrac{T_x+T_y}{2}\dfrac{x_i+y_i}{T_x\Big(1+\dfrac{T_y}{\sum_{j=1}^m x_j}\Big)}$	$\dfrac{T_x+T_y}{2}\dfrac{x_i}{T_x}$

このように、MLE が Asano の形と同じようになる例が、多項分布以外にもある。しかしながら、どのような場合に、MLE が Asano が示した形と同じになるのか？について、はっきとした見解を見出せていなかった。これが本研究に着手した動機である。

2. パラメーター変換によって MLE を求める

Proposition 1.1, 1.2, 1.3 いずれの場合においても、MLE を以下のようなパラメーター変換によって求めるのが容易である。

Proposition 1.1 (Asano) の場合 $u_i = \dfrac{p_i}{\sum_{j=0}^m p_j}$, $(i \le m)$, $t_0 = \sum_{j=0}^m p_j$, $t_i = p_i$, $(i > m)$ なるパラメータ変換を用いる。逆変換が $p_i = t_0 u_i$, $(i \le m)$ ゆえ

$$L \propto t_0^{x_0+\cdots+x_m} t_{m+1}^{x_{m+1}} \cdots t_k^{x_k} u_0^{y_0+x_0} \cdots u_m^{y_m+x_m}$$

である。よって $t_0 = \dfrac{\sum_{j=0}^m x_j}{N_1}$, $t_j = \dfrac{x_j}{N_1}$, $(j > m)$, $u_i = \dfrac{x_i+y_i}{\sum_{j=0}^m (x_j+y_j)}$. これより直ちに (1) を得る。

Proposition 1.2 の場合 $p_0 = 1-(p_1+\cdots+p_k)$ と定める。そして、パラメーター変換

$$t = p_0, \quad v = \frac{p_1+\cdots+p_m}{p_1+\cdots+p_m+p_{m+1}+\cdots+p_k}, \quad \xi_i = \frac{p_i}{p_1+\cdots+p_m}, \quad (i=1,2,\cdots,m)$$
$$\eta_j = \frac{p_j}{p_{m+1}+\cdots+p_k}, \quad (j=m+1,\cdots,k)$$

を用いる。逆変換が $p_0 = t,\ p_i = (1-t)v\xi_i \quad (i \le m),\ \ p_i = (1-t)(1-v)\eta_i \ \ (i > m)$ であるから $L \propto t^{r_1+r_2}(1-t)^{T_x+T_y} v^{\sum_{j=1}^m x_j}(1-v)^{\sum_{j=m+1}^k x_j} \xi_1^{x_1+y_1}\cdots\xi_m^{x_m+y_m}\eta_{m+1}^{x_{m+1}}\cdots\eta_k^{x_k}$. これより

$$\hat{t} = \frac{r_1+r_2}{T_x+T_y+r_1+r_2},\ \ \hat{v} = \frac{\sum_{j=1}^m x_j}{T_x},\ \ \hat{\xi}_i = \frac{x_i+y_i}{\sum_{j=1}^m (x_j+y_j)},\ (i \le m),$$
$$\hat{\eta}_i = \frac{x_i}{\sum_{j=m+1}^k x_j},\ (i > m)$$

を得る。$\hat{p}_0 = \hat{t},\ \ \hat{p}_i = (1-\hat{t})\hat{v}\hat{\xi}_i,\ \ (i \le m),\ \ \hat{p}_i = (1-\hat{t})(1-\hat{v})\hat{\eta}_i,\ \ (i > m)$ に代入することで (2) が得られる。

Proposition 1.3 の場合 便宜的に

$$\xi_i = (\lambda_1 + \cdots + \lambda_m + \cdots + \lambda_k)\frac{\lambda_i}{\lambda_1+\cdots+\lambda_m}$$

と置くと、$X_i \sim Poisson(\lambda_i),\ \ Y_j \sim Poisson(\xi_j)$ である。変数変換

$$s = \lambda_1 + \cdots + \lambda_m + \lambda_{m+1} + \cdots + \lambda_k,\quad t_0 = \frac{\lambda_1+\cdots+\lambda_m}{\lambda_1+\cdots+\lambda_m+\cdots+\lambda_k},$$
$$u_i = \frac{\lambda_i}{\lambda_1+\cdots+\lambda_m},\ \ (i = 1,2,\cdots,m),\quad t_j = \frac{\lambda_j}{\lambda_{m+1}+\cdots+\lambda_k},\ \ (j = m+1,\cdots,k)$$

を用いると、逆変換が $\lambda_i = st_0u_i,\ \ (i \le m),\ \ \lambda_i = s(1-t_0)t_i,\ \ (i > m),\ \xi_i = st_i$ であるから

$$L \propto s^{\sum_{i=1}^k x_i + \sum_{i=1}^m y_i}\exp(-2s)u_1^{x_1+y_1}\cdots u_m^{x_m+y_m} t_0^{\sum_{i=1}^m x_i}(1-t_0)^{\sum_{i=m+1}^k x_i} t_{m+1}^{x_{m+1}}\cdots t_k^{x_k}.$$

これより、

$$\hat{t}_0 = \frac{\sum_{j=1}^m x_j}{\sum_{j=1}^k x_j} = \frac{\sum_{j=1}^m x_j}{T_x},\quad \hat{t}_j = \frac{x_j}{\sum_{j=m+1}^k x_j},\quad j = m+1, m+2, \cdots, k,$$
$$\hat{u}_i = \frac{x_i+y_i}{\sum_{l=1}^m (x_l+y_l)},\quad i = 1,2,\cdots,m,\quad \hat{s} = \frac{T_x+T_y}{2}$$

が得られる。よって、$\hat{\lambda}_i = \hat{s}\hat{t}_0\hat{u}_i,\ \ (i \le m\ ),\ \ \hat{\lambda}_i = \hat{s}(1-\hat{t}_0)\hat{t}_i,\ \ (i > m\ )$ により、(3) が得られる。

3. Several extensions

本節では、 Proposition 1.3 をいくつかの状況に拡張することを考察する。

Proposition 3.1 l, m, k は $l < m < k$ なる自然数。

$$\xi_i = (\lambda_1 + \cdots + \lambda_l + \cdots + \lambda_m)\frac{\lambda_i}{\lambda_1+\cdots+\lambda_l},\qquad (i = 1,2,\cdots,l) \tag{4}$$

と定める。確率変数 $X_1,\cdots,X_k$, $Y_1,\cdots,Y_l$, $Y_{m+1},\cdots,Y_k$, は互いに独立で $X_i \sim Poisson(\lambda_i)$ for $i=1,\cdots,k$, $Y_i \sim Poisson(\xi_i)$ for $i=1,\cdots,l$, $Y_i \sim Poisson(\lambda_i)$ for $i=m+1,\cdots,k$, ならば、λ_i の MLE $\hat{\lambda}_i$ は

$$\hat{\lambda}_i = \begin{cases} \dfrac{T'_x+T'_y}{2}\dfrac{x_i+y_i}{T'_x\left(1+\dfrac{T'_y}{\sum_{i=1}^{l}x_i}\right)} & \text{if } i\le l, \\ \dfrac{T'_x+T'_y}{2}\dfrac{x_i}{T'_x} & \text{if } l<i\le m, \\ \dfrac{x_i+y_i}{2} & \text{if } i>m \end{cases} \tag{5}$$

である。但し $T'_x=\sum_{i=1}^{m}x_i$, $T'_y=\sum_{i=1}^{l}y_i$.

これは、パラメータ変換

$$\begin{aligned} &s=\lambda_1+\cdots+\lambda_m+\lambda_{m+1}+\cdots+\lambda_s, \qquad t=\frac{\lambda_1+\cdots+\lambda_m}{\lambda_1+\cdots+\lambda_m+\cdots+\lambda_k}, \\ &u=\frac{\lambda_1+\cdots+\lambda_l}{\lambda_1+\cdots+\lambda_m}, \qquad \theta_i=\frac{\lambda_i}{\lambda_1+\cdots+\lambda_l}, \ (i=1,2,\cdots,l), \\ &v_i=\frac{\lambda_i}{\lambda_{l+1}+\cdots+\lambda_m}, \ (j=l+1,\cdots,m), \\ &w_i=\frac{\lambda_i}{\lambda_{m+1}+\cdots+\lambda_k}, \ (j=m+1,\cdots,k) \end{aligned}$$

を用いれば容易に求められる。

Proposition 3.2 $l<m<k$ および ξ_i は、Proposition 2.1 と同じ。$X_1,\cdots,X_k$, $Y_1,\cdots,Y_l$, および Y_* は互いに独立で、$X_i \sim Poisson(\lambda_i)$, for all $i=1,\cdots,k$, $Y_i \sim Poisson(\xi_i)$, for $i=1,\cdots,l$, $Y_* \sim Poisson(\sum_{i=m+1}^{k}\lambda_i)$, とする。このとき λ_i の MLE は

$$\hat{\lambda}_i = \begin{cases} \dfrac{T'_x+T'_y}{2}\dfrac{x_i+y_i}{T'_x\left(1+\dfrac{T'_y}{\sum_{i=1}^{l}x_i}\right)}, & \text{if } i\le l, \\ \dfrac{T'_x+T'_y}{2}\dfrac{x_i}{T'_x}, & \text{if } l<i\le m, \\ \dfrac{1}{2}\dfrac{\sum_{i=m+1}^{k}x_i+y_*}{\sum_{i=m+1}^{k}x_i}x_i, & \text{if } i>m \end{cases} \tag{6}$$

である。

これも、Proposition 3.1 と同じ変数変換で求められる。

Proposition 3.3 $l<m<k$ は、Proposition 2.1 と同じ。$\{X_i\}_{i=1,\cdots,k}$ および $\{Y_i\}_{i=1,\cdots,l}$ は互いに独立で $X_i \sim Poisson(\lambda_i)$, $Y_i \sim Poisson(\xi_i)$, ここで

$$\xi_i=(\lambda_1+\cdots+\lambda_l+\cdots+\lambda_m)\frac{\lambda_i}{\lambda_1+\cdots+\lambda_l}, \qquad i=1,2,\cdots,l. \tag{7}$$

このとき、λ_i のMLEは

$$\hat{\lambda}_i = \begin{cases} \dfrac{T'_x + T_y}{2} \dfrac{x_i + y_i}{T'_x\Big(1 + \dfrac{T_y}{\sum_{i=1}^{l} x_i}\Big)}, & \text{if } i \le l, \\ \dfrac{T'_x + T_y}{2} \dfrac{x_i}{T'_x}, & \text{if } l < i \le m, \\ x_i, & \text{if } i > m \end{cases} \tag{8}$$

である。

Proposition 3.1 で用いたパラメータ変換を用いると、

$$L \propto s^{T_x+T_y} t^{(\Sigma_{i=1}^{m} x_i)+T_y} (1-t)^{\Sigma_{i=m+1}^{k} x_i} u^{\Sigma_{i=1}^{l} x_i} (1-u)^{\Sigma_{i=l+1}^{m} x_i} \times \prod_{i=1}^{l} \theta_i^{x_i+y_i} \prod_{i=l+1}^{m} v_i^{x_i} \prod_{i=m+1}^{k} w_i^{x_i} \times \exp(-s(1+t)).$$

これより直ちに

$$\hat{\theta}_i = \frac{x_i + y_i}{\sum_{i=1}^{l}(x_i + y_i)}, \quad (i = 1, 2, \cdots, l), \quad \hat{v}_i = \frac{x_i}{\sum_{i=l+1}^{m} x_i}, \quad (i = l+1, l+2, \cdots, m),$$

$$\hat{w}_i = \frac{x_i}{\sum_{i=m+1}^{k} x_i}, \quad (i = m+1, m+2, \cdots, k), \quad \hat{u} = \frac{\sum_{i=1}^{l} x_i}{\sum_{i=1}^{m} x_i}$$

を得る。s および t のMLEを、対数尤度

$$\log L = C + (T_x + T_y)\log s + ((\sum_{i=1}^{m} x_i) + T_y)\log t + (\sum_{i=m+1}^{k} x_i)\log(1-t) - s(1+t)$$

より求めてみる。尤度方程式は

$$0 = \frac{\partial L}{\partial s} = \frac{T_x + T_y}{s} - (1+t), \tag{9}$$

$$0 = \frac{\partial L}{\partial t} = \frac{\sum_{i=1}^{m} x_i + T_y}{t} - \frac{\sum_{i=m+1}^{k} x_i}{1-t} - s \tag{10}$$

であるから、(9) より $s = \dfrac{T_x + T_y}{1+t}$. を得る。これを (10) に代入することで

$$0 = \frac{\sum_{i=1}^{m} x_i + T_y}{t} - \frac{\sum_{i=m+1}^{k} x_i}{1-t} - \frac{T_x + T_y}{1+t} \tag{11}$$

が得られる。この式は

$$\begin{aligned} 0 &= (-\sum_{i=1}^{m} x_i - T_y - \sum_{i=m+1}^{k} x_i + T_x + T_y)t^2 - (\sum_{i=m+1}^{k} x_i + T_x + T_y)t + \sum_{i=1}^{m} x_i + T_y \\ &= (\sum_{i=m+1}^{k} x_i + T_x + T_y)t + \sum_{i=1}^{m} x_i + T_y \end{aligned}$$

と同値なので、$\hat{t}=\dfrac{\sum_{i=1}^{m}x_i+T_y}{\sum_{i=m+1}^{k}x_i+T_x+T_y}$ が得られる。これより

$$\hat{s}=\frac{T_x+T_y}{1+\hat{t}}=\frac{T_x+T_y}{1+\dfrac{\sum_{i=1}^{m}x_i+T_y}{\sum_{i=m+1}^{k}x_i+T_x+T_y}}=\frac{(T_x+T_y)(\sum_{i=m+1}^{k}x_i+T_x+T_y)}{T_x+T_y+\sum_{i=m+1}^{k}x_i+\sum_{i=1}^{m}x_i+T_y}$$
$$=\frac{T_x+T_y+\sum_{i=m+1}^{k}x_i}{2}$$

が得られ、あとは MLE の invariant property によって (8) が求められる。

Proposition 3.4 $\{X_i\}_{i=1,\cdots,k}$ および $\{Y_i\}_{i=1,\cdots,m}$ $(m<k)$ は互いに独立で、$X_i\sim Poisson(\lambda_i)$, $(i=1,\cdots,k)$, $Y_i\sim Poisson(\xi_i)$, $(i=1,\cdots l)$, $Y_i\sim Poisson(\eta_i)$, $(i=l+1,\cdots m)$ と仮定する。ここで

$$\xi_i=(\lambda_1+\cdots+\lambda_l+\cdots+\lambda_m)\frac{\lambda_i}{\lambda_1+\cdots+\lambda_l},\qquad (i=1,2,\cdots,l),$$
$$\eta_i=(\lambda_{m+1}+\cdots+\lambda_k)\frac{\lambda_i}{\lambda_{l+1}+\cdots+\lambda_m},\qquad (i=l+1,l+2,\cdots,m).$$

そうすると、λ_i の MLE は

$$\frac{T_x'+T_y'}{2}\frac{x_i+y_i}{T_x'\Big(1+\dfrac{T_y'}{\sum_{i=1}^{l}x_i}\Big)},\ (i\le l),\qquad \frac{T_x'+T_y'}{2}\frac{x_i+y_i}{T_x'\Big(1+\dfrac{\sum_{i=l+1}^{m}y_i}{\sum_{i=l+1}^{m}x_i}\Big)},\ (l<i\le m),$$
$$\Big\{1+\frac{\sum_{i=l+1}^{m}y_i}{\sum_{i=m+1}^{k}x_i}\Big\}\frac{x_i}{2},\quad (i>m)$$

である。なお、$T_y'=\sum_{i=1}^{l}y_i$.

これも、Proposition 3.1 と同じ変数変換で求められる。

4. Related multinomial models

前章の結果は、多項分布の場合においてもほぼ同様に成り立つ。つまり、Prop. 1.3 の拡張として Prop 3.1 から Prop 3.4 を得たのであるが、ほぼ同様に Prop 1.1 を拡張することができる。

Proposition 4.1 l,k,m は $m<k$ なる自然数。確率変数 $\mathbf{X}$, $\mathbf{Y}$ は互いに独立で

$$\mathbf{X}=(X_0,\cdots,X_l,\cdots,X_m,\cdots,X_k)\sim \text{Multinomial}(N_1;p_0,\cdots,p_k),$$
$$\mathbf{Y}=(Y_0,\cdots,Y_l,Y_{m+1},\cdots,Y_k)$$
$$\sim \text{Multinomial}(N_2;\sum_{j=0}^{m}p_j\frac{p_0}{\sum_{j=0}^{l}p_j},\cdots,\sum_{j=0}^{m}p_j\frac{p_l}{\sum_{j=0}^{l}p_j},p_{m+1},\cdots,p_k),$$

とすると、$\hat{p}_i(\mathbf{x},\mathbf{y})$ は

$$\hat{p}_i(\mathbf{x},\mathbf{y}) = \begin{cases} \dfrac{N_1'+N_2'}{N_1+N_2}\dfrac{x_i+y_i}{N_1'\Big(1+\dfrac{N_2'}{\sum_{i=0}^{l}x_i}\Big)}, & \text{if} \quad i \le l, \\ \dfrac{N_1'+N_2'}{N_1+N_2}\dfrac{x_i}{N_1'}, & \text{if} \quad l < i \le m, \\ \dfrac{x_i+y_i}{N_1+N_2}, & \text{if} \quad i > m \end{cases} \tag{12}$$

である。但し、$N_1' = \sum_{i=0}^{m} x_i, \quad N_2' = \sum_{i=0}^{l} y_i.$

Proposition 4.2 l,k,m は $m<k$ なる自然数。確率変数 $\mathbf{X}, \mathbf{Y}$ は互いに独立で

$$\mathbf{X} = (X_0,\cdots,X_l,\cdots,X_m,\cdots,X_k) \sim \text{Multinomial}(N_1; p_0,\cdots,p_k),$$
$$\mathbf{Y} = (Y_0,\cdots,Y_l,Y_*)$$
$$\sim \text{Multinomial}(N_2; \sum_{j=0}^{m} p_j\frac{p_0}{\sum_{j=0}^{l}p_j},\cdots,\sum_{j=0}^{m} p_j\frac{p_l}{\sum_{j=0}^{l}p_j}, \sum_{i=m+1}^{k} p_i)$$

ならば、$\hat{p}_i(\mathbf{x},\mathbf{y})$ は

$$\hat{p}_i(\mathbf{x},\mathbf{y}) = \begin{cases} \dfrac{N_1'+N_2'}{N_1+N_2}\dfrac{x_i+y_i}{N_1'\Big(1+\dfrac{N_2'}{\sum_{i=0}^{l}x_i}\Big)}, & \text{if} \quad i \le l, \\ \dfrac{N_1'+N_2'}{N_1+N_2}\dfrac{x_i}{N_1'}, & \text{if} \quad l < i \le m, \\ \dfrac{\sum_{i=m+1}^{k}x_i+y_*}{N_1+N_2}\dfrac{x_i}{\sum_{i=m+1}^{k}x_i}, & \text{if} \quad i > m \end{cases} \tag{13}$$

と書ける。

Proposition 4.3 l,k,m は $m<k$ なる自然数。確率変数 $\mathbf{X}, \mathbf{Y}$ は互いに独立で

$$\mathbf{X} = (X_0,\cdots,X_l,\cdots,X_m,\cdots,X_k) \sim \text{Multinomial}(N_1; p_0,\cdots,p_k),$$
$$\mathbf{Y} = (Y_0,\cdots,Y_l,Y_{l+1},\cdots,Y_m)$$
$$\sim \text{Multinomial}(N_2; \sum_{j=0}^{m} p_j\frac{p_0}{\sum_{j=0}^{l}p_j},\cdots,\sum_{j=0}^{m} p_j\frac{p_l}{\sum_{j=0}^{l}p_j},$$
$$\sum_{j=m+1}^{k} p_j\frac{p_{l+1}}{\sum_{j=l+1}^{m}p_j},\cdots,\sum_{j=m+1}^{k} p_j\frac{p_m}{\sum_{j=l+1}^{m}p_j})$$

ならば、p_i は

$$\hat{p}_i(\mathbf{x},\mathbf{y})=\begin{cases}\dfrac{N_1'+N_2'}{N_1+N_2}\dfrac{x_i+y_i}{N_1'\Big(1+\dfrac{N_2'}{\sum_{i=0}^{l}x_i}\Big)}, & \text{if}\quad i\le l,\\[2ex] \dfrac{N_1'+N_2'}{N_1+N_2}\dfrac{x_i+y_i}{N_1'\Big(1+\dfrac{\sum_{i=l+1}^{m}y_i}{\sum_{i=l+1}^{m}x_i}\Big)}, & \text{if}\quad l<i\le m,\\[2ex] \Big\{1+\dfrac{\sum_{i=l+1}^{m}y_i}{\sum_{i=m+1}^{k}x_i}\Big\}\dfrac{x_i}{N_1+N_2}, & \text{if}\quad i>m\end{cases}\tag{14}$$

である。但し、$N_1'=\sum_{i=0}^{m}x_i,\quad N_2'=\sum_{i=0}^{l}y_i.$

Proposition 4.1, 4.2 および 4.3 いずれも、パラメーター変数変換

$$t=p_0+\cdots+p_m,\quad u=\frac{p_0+\cdots+p_l}{p_0+\cdots+p_l+\cdots+p_m},\quad \theta_i=\frac{p_i}{p_0+\cdots+p_l}\ (i\le l),$$
$$v_i=\frac{p_i}{p_{l+1}+\cdots+p_m}\ (l<i\le m),\quad w_i=\frac{p_i}{p_{m+1}+\cdots+p_k}\ (m<i).$$

を用いれば容易。

5. パラメータが樹木構造でない場合

Proposition 1.1 ～ 1.3 は、パラメータが樹木構造で書けている。これに対して、Proposition 2.1 ～ 2.5 は、パラメータが「完全に樹木構造」とは言えないものの、MLE が exact に求められたものである。そこで、パラメータが樹木構造ではないために MLE が exact には求めらないが、EM アルゴリズムを用いて MLE が求められるようなモデルについて議論する。

Problem $\{X_i\}_{i=1,\cdots,k}$, $\{Y_i\}_{i=1,\cdots,m}$, $\{Z_i\}_{i=l,\cdots,k}$, および $\{W_i\}_{i=1,\cdots,l,m+1,\cdots,k}$ は。互いに独立で、$X_i\sim Poisson(\lambda_i)$, $Y_i\sim Poisson(\mu_i)$, $Z_i\sim Poisson(\eta_i)$, $W_i\sim Poisson(\xi_i)$, とする。但し、

$$\mu_i=(\lambda_1+\cdots+\lambda_m)\frac{\lambda_i}{\lambda_1+\cdots+\lambda_l}\qquad i=1,\cdots,l,$$
$$\eta_i=(\lambda_l+\cdots+\lambda_k)\frac{\lambda_i}{\lambda_{l+1}+\cdots+\lambda_m}\qquad i=l+1,\cdots,m,$$
$$\xi_i=(\lambda_1+\cdots+\lambda_l+\lambda_{m+1}+\cdots+\lambda_k)\frac{\lambda_i}{\lambda_{m+1}+\cdots+\lambda_k}\qquad i=m+1,\cdots,k.$$

このとき、λ_i の MLE を求めたい。

$$s_1=\sum_{i=1}^{l}\lambda_i,\quad s_2=\sum_{i=l+1}^{m}\lambda_i,\quad s_3=\sum_{i=m+1}^{k}\lambda_i,\quad t_i=\frac{\lambda_i}{\sum_{i=1}^{l}\lambda_i}\ (i\le l)$$
$$u_i=\frac{\lambda_i}{\sum_{i=l+1}^{m}\lambda_i}\ (l<i\le m),\quad v_i=\frac{\lambda_i}{\sum_{i=m+1}^{k}\lambda_i}\ (i\ge m)$$

なるパラメータ変換を用いる。上記の逆変換は

$$\lambda_i = s_1 t_i,\ (i \le l), \quad \lambda_i = s_2 u_i,\ (l < i \le m), \quad \lambda_i = s_3 v_i,\ (i \ge m),$$
$$\mu_i = (s_1+s_2) t_i,\ (i \le l), \quad \eta_i = (s_2+s_3) u_i,\ (l < i \le m), \quad \xi_i = (s_1+s_3) v_i,\ (i \ge m).$$

これを用いて

$$L = \Big(\prod_{i=1}^{k} \frac{\lambda_i^{x_i}}{x_i!}\exp(-\lambda_i)\Big)\Big(\prod_{i=1}^{l} \frac{\mu_i^{y_i}}{y_i!}\exp(-\mu_i)\Big)\Big(\prod_{i=l+1}^{m} \frac{\eta_i^{z_i}}{z_i!}\exp(-\eta_i)\Big)\Big(\prod_{i=m+1}^{k} \frac{\xi_i^{w_i}}{w_i!}\exp(-\xi_i)\Big)$$
$$\propto s_1^{\sum_{i=1}^{l} x_i} s_2^{\sum_{i=l+1}^{m} x_i} s_3^{\sum_{i=m+1}^{k} x_i} (s_1+s_2)^{\sum_{i=1}^{l} y_i} (s_2+s_3)^{\sum_{i=l+1}^{m} z_i} (s_1+s_3)^{\sum_{i=m+1}^{k} w_i}$$
$$\times\Big(\prod_{i=1}^{l} t_i^{x_i+y_i}\Big)\Big(\prod_{i=l+1}^{m} u_i^{x_i+z_i}\Big)\Big(\prod_{i=m+1}^{k} v_i^{x_i+w_i}\Big)\exp(-3(s_1+s_2+s_3)).$$

これより直ちに

$$\hat{t}_i = \frac{x_i+y_i}{\sum_{i=1}^{l}(x_i+y_i)}, \quad \hat{u}_i = \frac{x_i+z_i}{\sum_{i=l+1}^{m}(x_i+z_i)}, \quad \hat{v}_i = \frac{x_i+w_i}{\sum_{i=m+1}^{k}(x_i+w_i)}$$

を得る。さて、$\hat{s}_1$, $\hat{s}_2$, $\hat{s}_3$ は、次のようにして求められる：

$s = s_1+s_2+s_3$, $\theta_i = \dfrac{s_i}{s_1+s_2+s_3}$ と変数変換すると、$s_1 = s\theta_1$, $s_2 = s\theta_2$, $s_3 = s\theta_3$.
そして、表記の簡略化のため

$$T(x) = \sum_{i=1}^{k} x_i, \quad T(y) = \sum_{i=1}^{l} y_i, \quad T(z) = \sum_{i=l+1}^{m} z_i, \quad T(w) = \sum_{i=m+1}^{k} w_i$$

と定めると、

$$L \propto (s\theta_1)^{\sum_{i=1}^{l} x_i} (s\theta_2)^{\sum_{i=l+1}^{m} x_i} (s\theta_3)^{\sum_{i=m+1}^{k} x_i} (s(\theta_1+\theta_2))^{T(y)} (s(\theta_2+\theta_3))^{T(z)}$$
$$\times (s(\theta_1+\theta_3))^{T(w)} \exp(-3s)$$
$$= s^{T(x)}\exp(-3s) \times \theta_1^{\sum_{i=1}^{l} x_i} \theta_2^{\sum_{i=l+1}^{m} x_i} \theta_3^{\sum_{i=m+1}^{k} x_i} (\theta_1+\theta_2)^{T(y)} (\theta_2+\theta_3)^{T(z)} (\theta_1+\theta_3)^{T(w)}.$$

これより $\hat{s} = \dfrac{T(x)+T(y)+T(z)+T(w)}{3}$. さて、

$$L(\theta_1,\theta_2,\theta_3) \propto \theta_1^{\sum_{i=1}^{l} x_i} \theta_2^{\sum_{i=l+1}^{m} x_i} \theta_3^{\sum_{i=m+1}^{k} x_i} \times (\theta_1+\theta_2)^{T(y)} (\theta_2+\theta_3)^{T(z)} (\theta_1+\theta_3)^{T(w)}$$

であるから、$\hat{\theta}_i$, $(i=1,2,3)$ を EM アルゴリズムによって次のように求めることができる。最初に、初期値 $\theta_i^{(0)}$ $(i=1,2,3)$ を

$$\theta_1^{(0)} = \frac{\sum_{i=1}^{l} x_i + T(y)/2 + T(w)/2}{T_x+T_y+T_z+T_w}, \quad \theta_2^{(0)} = \frac{\sum_{i=l+1}^{m} x_i + T(y)/2 + T(z)/2}{T_x+T_y+T_z+T_w},$$
$$\theta_3^{(0)} = \frac{\sum_{i=m+1}^{k} x_i + T(z)/2 + T(w)/2}{T_x+T_y+T_z+T_w}$$

に選ぶ。各 $k = 0, 1, \cdots,$ に対して、$T^{(k)}(y)[1]$, $T^{(k)}(y)[2]$, $T^{(k)}(z)[2]$, $T^{(k)}(z)[3]$, $T^{(k)}(w)[1]$, $T^{(k)}(w)[3]$, (E-step) および $\theta_1^{(k+1)}$, $\theta_2^{(k+1)}$, $\theta_3^{(k+1)}$, (M-step) を次のように構築する。

E-step ：完全データの条件付期待値を計算するステップ

$$T^{(k)}(y)[i] = E(T(y)[i] \mid T(y); \theta_1^{(k)}, \theta_2^{(k)}) = T(y) \frac{\theta_i^{(k)}}{\theta_1^{(k)} + \theta_2^{(k)}}, \quad i = 1, 2,$$

$$T^{(k)}(z)[i] = E(T(z)[i] \mid T(z); \theta_2^{(k)}, \theta_3^{(k)}) = T(z) \frac{\theta_i^{(k)}}{\theta_2^{(k)} + \theta_3^{(k)}}, \quad i = 2, 3,$$

$$T^{(k)}(w)[i] = E(T(w)[1] \mid T(w); \theta_1^{(k)}, \theta_3^{(k)}) = T(w) \frac{\theta_i^{(k)}}{\theta_1^{(k)} + \theta_3^{(k)}}, \quad i = 1, 3.$$

M-step ：完全データ (の近似値) を使って、パラメータの値を更新するステップ。すなわち、データ $\sum_{i=1}^{l} x_i + T^{(k)}(y)[1] + T^{(k)}(w)[1]$, $\sum_{i=l+1}^{m} x_i + T^{(k)}(y)[2] + T^{(k)}(z)[2]$, $\sum_{i=m+1}^{k} x_i + T^{(k)}(z)[3] + T^{(k)}(w)[3]$ が与えられたときの尤度を最大にする $\theta_1^{(k+1)}$, $\theta_2^{(k+1)}$, $\theta_3^{(k+1)}$ の値は

$$\theta_1^{(k+1)} = \frac{\sum_{i=1}^{l} x_i + T^{(k)}(y)[1] + T^{(k)}(w)[1]}{T(x) + T(y) + T(z) + T(w)},$$

$$\theta_2^{(k+1)} = \frac{\sum_{i=l+1}^{m} x_i + T^{(k)}(y)[2] + T^{(k)}(z)[2]}{T(x) + T(y) + T(z) + T(w)},$$

$$\theta_3^{(k+1)} = \frac{\sum_{i=m+1}^{k} x_i + T^{(k)}(z)[3] + T^{(k)}(w)[3]}{T(x) + T(y) + T(z) + T(w)}$$

である。各 $i = 1, 2, 3$ に対して、EM アルゴリズムにより $\lim_{k\to\infty} \theta_i^{(k)} = \hat{\theta}_i$ である。MLE の invariance property により、

$$\hat{\lambda}_i = \hat{s}_1 \hat{\theta}_i \hat{t}_i \ \ (i \le l), \quad \hat{\lambda}_i = \hat{s}_2 \hat{\theta}_i \hat{u}_i \ \ (l < i \le m), \quad \hat{\lambda}_i = \hat{s}_3 \hat{\theta}_i \hat{v}_i \ \ (i > m).$$

を得る。 下記の表は、E-step, M-step における 観測された値および完全データと確率との対応を示したものである

確率	観測された値
θ_1	$\sum_{i=1}^{l} x_i$
$\theta_1 + \theta_2$	$T(y)$
θ_2	$\sum_{i=l+1}^{m} x_i$
$\theta_2 + \theta_3$	$T(z)$
θ_3	$\sum_{i=m+1}^{k} x_i$
$\theta_3 + \theta_1$	$T(w)$

セル確率	完全データ
θ_1	$\sum_{i=1}^{l} x_i + T(y)[1] + T(w)[1]$
θ_2	$\sum_{i=l+1}^{m} x_i + T(y)[2] + T(z)[2]$
θ_3	$\sum_{i=m+1}^{k} x_i + T(z)[3] + T(w)[3]$

Acknowledgements

本研究は、科学研究費 基盤研究 (C) 課題番号 20500261 の研究成果の一部である。この科学研究費の援助を、記して感謝申し上げます。

References

[1] Asano, C. (1965) On estimating multinomial probabilities by pooling incomplete samples, *Annals of the Institute of Statistical Mathematics* **17**, 1-13.

[2] Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society* **B 39**, 1-38.

[3] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1997). *Discrete Multivariate Distributions.* Wiley.

[4] McLachlan, G. J., and Krishnan, T. ((1977) *The EM Algorithm and Extensions.* Wiley.