多項分布のパラメータ推定

AUTHOR: Bunji Square

1 から m までの目を持つサイコロを想定する。このサイコロの 各目の出る確率 $\theta_1 ... \theta_m$ が未知だとする ($0 > \theta_j > 1$ & $\Sigma_{j=1}^m \theta_j = 1$)。 サイコロを n 回投げた結果が、$\boldsymbol{x} = \boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_i}, ..., \boldsymbol{x_n}$ で i 回目の試行で j 面が出ることを $x_{ij} = 1$ 出ないときは $x_{ij} = 0$ として、i 回目の試行を $\boldsymbol{x_i} = x_{i1}, ..., x_{ij}, ... x_{im}$ と表現する。

$$x_{ij} \in \{0, 1\} \quad \Sigma_{j=1}^m x_{ij} = 1$$

さて n 回の試行のうち、各面の出た回数は r_j 回だったとする。

$$r_j = \Sigma_{i=1}^n x_{ij} \ \ (0 \leq j \leq n) \quad n = \Sigma_{j=1}^m r_j$$

この結果をもとに、確率ベクトル $\boldsymbol{\theta} = \theta_1, ..., \theta_m$ を推定してみる。

例として通常の 6 面を持つサイコロを三回振ることを考える。 1回目の目が3、2回目の目が1、3回目の目が6とすると、 これらの確率変数 $\boldsymbol{x} = \boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}$ は次のように表現されることとなる。

試行回数	出た目	確率変数	
1回目	3	$\boldsymbol{x_1}$	$x_{11} = 0, x_{12} = 0, x_{13} = 1, x_{14} = 0, x_{15} = 0, x_{16} = 0$
2回目	1	$\boldsymbol{x_2}$	$x_{21} = 1, x_{22} = 0, x_{23} = 0, x_{24} = 0, x_{25} = 0, x_{26} = 0$
3回目	6	$\boldsymbol{x_3}$	$x_{31} = 0, x_{32} = 0, x_{33} = 0, x_{34} = 0, x_{35} = 0, x_{36} = 1$

1回目の試行の確立は次のように表される。

$$p(\boldsymbol{x_1}|\boldsymbol{\theta}) = \theta_1^{x_{11}} \theta_2^{x_{12}} \theta_3^{x_{13}} \theta_4^{x_{14}} \theta_5^{x_{15}} \theta_6^{x_{16}} = \theta_1^0 \theta_2^0 \theta_3^1 \theta_4^0 \theta_5^0 \theta_6^0 = \theta_3$$

同様に、2回目・3回目は次のようになる。

$$p(\boldsymbol{x_2}|\boldsymbol{\theta}) = \theta_1 p(\boldsymbol{x_3}|\boldsymbol{\theta}) = \theta_6$$

各試行はそれぞれ独立であることから、これらの結果の確率は次のようになる。

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = p(\boldsymbol{x_1}|\boldsymbol{\theta}) p(\boldsymbol{x_2}|\boldsymbol{\theta}) p(\boldsymbol{x_3}|\boldsymbol{\theta}) = \theta_3 \theta_1 \theta_6$$

少し一般化しておく。 i 回目の試行の確立は次のように表される。

$$p(\boldsymbol{x_i}|\boldsymbol{\theta}) = \theta_1^{x_{i1}} \cdots \theta_m^{x_{im}} = \Pi_{j=1}^m \theta_j^{x_{ij}}$$

従って n 回の試行結果の確率 $\boldsymbol{x}$ は次のようになる。

$$\begin{aligned} p(\boldsymbol{x}|\boldsymbol{\theta}) &= \Pi_{i=1}^n p(\boldsymbol{x_i}|\boldsymbol{\theta}) \\ &= \Pi_{i=1}^n \Pi_{j=1}^m \theta_j^{x_{ij}} \end{aligned}$$

そして n 回のうち、各面の出た回数は r_i 回だったことから次のように変形できる。

$$\begin{aligned} p(\boldsymbol{x}|\boldsymbol{\theta}) &= \Pi_{i=1}^n \Pi_{j=1}^m \theta_j^{x_{ij}} \\ &= \Pi_{j=1}^m \theta_j^{\Sigma_{i=1}^n x_{ij}} \\ &= \Pi_{j=1}^m \theta_j^{r_j} \end{aligned}$$

このような試行結果 $\boldsymbol{x}$ の確率 $p(\boldsymbol{x}|\boldsymbol{\theta})$ を $\boldsymbol{x}$ の尤度関数と呼ぶ。

なお今回のサイコロの例では、$\theta_m = 1 - \Sigma_{j=1}^{m-1}\theta_j$ また $r_m = n - \Sigma_{j=1}^{m-1} r_j$ であることから 上記は m 次関数ではなく実質的に m-1 次関数であることも付け加えておく。

$$\begin{aligned} p(\boldsymbol{x}|\boldsymbol{\theta}) &= \Pi_{j=1}^{m-1} \theta_j^{r_j} \cdot \theta_m^{r_m} \\ &= \Pi_{j=1}^{m-1} \theta_j^{r_j} \cdot (1 - \Sigma_{j=1}^{m-1}\theta_j)^{n - \Sigma_{j=1}^{m-1} r_j} \end{aligned}$$

$\boldsymbol{\theta}$ の最尤推定値

$\boldsymbol{x}$ の尤度関数 $p(\boldsymbol{x}|\boldsymbol{\theta})$ を $\boldsymbol{\theta}$ の関数とみなして、$\boldsymbol{\theta}$　の最尤推定値を計算する。

$$L(\boldsymbol{\theta}) = p(\boldsymbol{x}|\boldsymbol{\theta}) = \Pi_{j=1}^m \theta_j^{r_j}$$

$$\log L(\boldsymbol{\theta}) = \Sigma_{j=1}^m r_j \log\theta_j = \Sigma_{j=1}^{m-1} r_j \log\theta_j + (n - \Sigma_{j=1}^{m-1} r_j)\log(1 - \Sigma_{j=1}^{m-1}\theta_j)$$

$$\frac{\partial}{\partial\theta_j}\log L(\boldsymbol{\theta}) = \frac{r_j}{\theta_j} - \frac{n - \Sigma_{j=1}^{m-1} r_j}{1 - \Sigma_{j=1}^{m-1}\theta_j} = \frac{r_j}{\theta_j} - \frac{r_m}{\theta_m} = 0$$

$$\theta_j = \frac{r_j}{r_m}\theta_m$$

θ_j はすべて足し合わせると 1 になるので、

$$\Sigma_{j=1}^m \theta_j = \Sigma_{j=1}^m \frac{r_j}{r_m}\theta_m = 1$$

$$\theta_m = \frac{r_m}{n} \quad \because \Sigma_{j=1}^m r_j = n$$

$$\therefore \hat{\theta_j}_{MLE} = \frac{r_j}{n}$$

つまり、ここでの最尤推定値とは出た目の回数の全体の試行回数に対する割合に相当する。

続いて、MAP 推定値と EAP 推定値を計算するために $\boldsymbol{\theta}$ の事後分布を考える。 多項分布の共益事前分布はディリクレ分布 (Dirichlet distribution)を使う。 $\boldsymbol{\alpha} = \alpha_1, \ldots \alpha_K$ をパラメータのベクトル、 $\boldsymbol{\theta} = \theta_1, \ldots \theta_K$ を確率ベクトルとする $K-1$ 次のディリクレ分布の関数密度関数は次式で表される。

$$Dir(\boldsymbol{\theta}; \boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \Pi_{i=1}^{K} \theta_i^{\alpha_i - 1}$$

ただし以下を満たすものとする。

$$\Sigma_{i=1}^{K} \theta_i = 1 \quad \theta_1, \ldots, \theta_K \geq 0 B(\boldsymbol{\alpha}) = \frac{\Pi_{i=1}^{K} \Gamma(\alpha_i)}{\Gamma(\Sigma_{i=1}^{K} \alpha_i)} = \frac{(\alpha_1 - 1)!(\alpha_2 - 1)! \ldots (\alpha_K - 1)!}{(\alpha_1 + \alpha_2 + \ldots + \alpha_K - 1)!}$$

$\theta_K = 1 - (\theta_1 + \ldots + \theta_{K-1})$ より、上の関数の次数は $K-1$ となる。 今回の場合は $K = m$ なので事前分布 $p(\boldsymbol{\theta})$は次の通り。

$$p(\boldsymbol{\theta}) = Dir(\boldsymbol{\theta}; \boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \Pi_{j=1}^{m} \theta_j^{\alpha_j - 1}$$

[積分公式]

$$\int_0^{\infty} \theta_1^{\alpha_1 - 1} \ldots \theta_n^{\alpha_n - 1} d\theta_1 \ldots d\theta_{n-1} = \frac{\Gamma(\alpha_1) \ldots \Gamma(\alpha_n)}{\Gamma(\alpha_1 + \ldots + \alpha_n)} \quad (1)$$

$$\theta_1 + \ldots + \theta_n = 1$$

$$\theta_n = 1 - \theta_1 - \theta_2 \ldots - \theta_{n-1}$$

$$\theta_1 + \ldots + \theta_{n-1} \leq 1$$

ディリクレ分布における θ_j の期待値を求めておく。

$$\begin{aligned}
E[\theta_1] &= \int_0^\infty \theta_1 Dir(\boldsymbol{\theta};\boldsymbol{\alpha})d\theta_1\ldots d\theta_{n-1} \\
&= \int_0^\infty \theta_1 \cdot \frac{1}{B(\boldsymbol{\alpha})}\Pi_{i=1}^n \theta_i^{\alpha_i-1} d\theta_1\ldots d\theta_{n-1} \\
&= \frac{1}{B(\boldsymbol{\alpha})}\int_0^\infty \theta_1^{\alpha}\theta_2^{\alpha_2-1}\ldots\theta_n^{\alpha_n-1}d\theta_1\ldots d\theta_{n-1} \\
&= \frac{1}{B(\boldsymbol{\alpha})}\frac{\Gamma(\alpha_1+1)\ldots\Gamma(\alpha_n)}{\Gamma(\Sigma_{i=1}^n\alpha_i+1)} & \text{(1) より} \\
&= \frac{\Gamma(\Sigma_{i=1}^n\alpha_i)}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_n)}\frac{\alpha_1\Gamma(\alpha_1)\ldots\Gamma(\alpha_n)}{(\Sigma_{i=1}^n\alpha_i)\Gamma(\Sigma_{i=1}^n\alpha_i)} \\
&= \frac{\alpha_1}{\Sigma_{i=1}^n\alpha_i}
\end{aligned}$$

従って、$\theta_j\ \ (j=1,\ldots,K-1)$ の期待値は次のようになる。

$$E[\theta_j] = \frac{\alpha_j}{\Sigma_{i=1}^n\alpha_i} \quad (j=1,\ldots,K-1)$$

参考： https://to-kei.net/distribution/dirichlet-distribution/dirichlet-distribution-derivation/

$p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ は次のようになる。

$$\begin{aligned}
p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}) &= \Pi_{j=1}^m \theta_j^{r_j}\cdot\frac{1}{B(\boldsymbol{\alpha})}\Pi_{j=1}^m\theta_j^{\alpha_j-1} \\
&= \frac{1}{B(\boldsymbol{\alpha})}\Pi_{j=1}^m\theta_j^{r_j+\alpha_j-1}
\end{aligned}$$

$p(\boldsymbol{x})$ は次のようになる。

$$\begin{aligned}
p(\boldsymbol{x}) &= \int_0^1 p(\boldsymbol{x}|\theta)p(\theta)d\theta \\
&= \int_0^1 \frac{1}{B(\boldsymbol{\alpha})}\Pi_{j=1}^m\theta_j^{r_j+\alpha_j-1}d\theta \\
&= \frac{1}{B(\boldsymbol{\alpha})}\int_0^1 \Pi_{j=1}^m\theta_j^{r_j+\alpha_j-1}d\theta \\
&= \frac{1}{B(\boldsymbol{\alpha})}\frac{\Pi_{j=1}^m\Gamma(r_j+\alpha_j)}{\Sigma_{j=1}^m(r_j+\alpha_j)}
\end{aligned}$$

以上から $p(\boldsymbol{\theta}|\boldsymbol{x})$ を求めると、次に示すようにディリクレ分布関数となる。

$$
\begin{aligned}
p(\boldsymbol{\theta}|\boldsymbol{x}) &= \frac{p(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{x})}\cdot p(\boldsymbol{\theta}) \\
&= \frac{1}{B(\boldsymbol{\alpha})}\Pi_{j=1}^{m}\theta_j^{r_j+\alpha_j-1}B(\boldsymbol{\alpha})\frac{\Sigma_{j=1}^{m}(r_j+\alpha_j)}{\Pi_{j=1}^{m}\Gamma(r_j+\alpha_j)} \\
&= \frac{\Sigma_{j=1}^{m}(r_j+\alpha_j)}{\Pi_{j=1}^{m}\Gamma(r_j+\alpha_j)}\Pi_{j=1}^{m}\theta_j^{r_j+\alpha_j-1} \\
&= \frac{\Sigma_{j=1}^{m}(\alpha'_j)}{\Pi_{j=1}^{m}\Gamma(\alpha'_j)}\Pi_{j=1}^{m}\theta_j^{\alpha'_j-1} & \because \alpha'_j = \alpha_j + r_j \\
&= \frac{1}{B(\boldsymbol{\alpha'})}\Pi_{j=1}^{m}\theta_j^{\alpha'_j-1} & \because \boldsymbol{\alpha}' = \alpha'_1,\ldots\alpha'_m \\
&= Dir(\boldsymbol{\theta};\boldsymbol{\alpha}')
\end{aligned}
$$

MAP 推定値。

$$
L(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{\Sigma_{j=1}^{m}(r_j+\alpha_j)}{\Pi_{j=1}^{m}\Gamma(r_j+\alpha_j)}\Pi_{j=1}^{m}\theta_j^{r_j+\alpha_j-1}
$$

$$
\begin{aligned}
\log L(\boldsymbol{\theta}) &= \Sigma_{j=1}^{m}(r_j+\alpha_j-1)\log\theta_j \\
&= \Sigma_{j=1}^{m}(r_j+\alpha_j-1)\log\theta_j + (r_m+\alpha_m-1)\log\theta_m \\
&= \Sigma_{j=1}^{m}(r_j+\alpha_j-1)\log\theta_j + (n-\Sigma_{j=1}^{m}r_j+\alpha_m-1)\log(1-\Sigma_{j=1}^{m}\theta_j)
\end{aligned}
$$

$$
\frac{\partial}{\partial\theta_j}\log L(\boldsymbol{\theta}) = \frac{r_j+\alpha_j-1}{\theta_j} - \frac{n-\Sigma_{j=1}^{m}r_j+\alpha_m-1}{1-\Sigma_{j=1}^{m}\theta_j} = 0
$$

$$
\theta_j = \frac{r_j+\alpha_j-1}{r_m+\alpha_m-1}\theta_m
$$

$$
\Sigma_{j=1}^{m}\theta_j = \frac{\theta_m}{r_m+\alpha_m-1}\Sigma_{j=1}^{m}(r_j+\alpha_j-1) = \frac{\theta_m}{r_m+\alpha_m-1}(n+\Sigma_{j=1}^{m}\alpha_j-m) = 1
$$

$$
\theta_m = \frac{r_m+\alpha_m-1}{n+\Sigma_{j=1}^{m}\alpha_j-m}
$$

$$
\therefore \hat{\theta_j}_{MAP} = \frac{r_j+\alpha_j-1}{n+\Sigma_{j=1}^{m}\alpha_j-m}
$$

事後分布がディリクレとなることから θ_j のEAP 推定値は次のようになる。

$$
\begin{aligned}
\hat{\theta_j}_{EAP} = E[\theta_j] &= \frac{\alpha'_j}{\Sigma_{j=1}^m \alpha'_j} \\
&= \frac{\alpha_j + r_j}{\Sigma_{j=1}^m (\alpha_j + r_j)} \\
&= \frac{r_j + \alpha_j}{n + \Sigma_{j=1}^m (\alpha_j)}
\end{aligned}
$$

従って、事前分布が一様分布（$\alpha_j = 1$）とすると θ_j の推定値 (最尤推定値・MAP 推定値・EAP 推定値) はそれぞれ次のようになる。

$$
\hat{\theta_j}_{MLE} = \hat{\theta_j}_{MAP} = \frac{r_j}{n}, \quad \hat{\theta_j}_{EAP} = \frac{r_j + 1}{n + m}
$$