ディリクレ分布の正規化項の求め方

正田 備也

平成 18 年 8 月 4 日

ディリクレ分布とは、確率分布の確率分布として使うことのできる、確率分布である。例えば、単語 $w_j, j = 1, \dots, m$ が出現する確率を θ_j とすると、次のようにディリクレ分布を用いて、単語の出現確率分布の確率分布を表すことができる。

$$P(\theta) = \frac{\Gamma(\sum_j \alpha_j)}{\prod_j \Gamma(\alpha_j)} \prod_j \theta_j^{\alpha_j - 1} \tag{1}$$

$\frac{\Gamma(\sum_j \alpha_j)}{\prod_j \Gamma(\alpha_j)}$ は正規化項だが、この求め方を以下に示す。

ディリクレ分布の $\prod_j \theta_j^{\alpha_j - 1}$ という部分について、$\{\theta | \sum_j \theta_j = 1, \theta_j \geq 0\}$ という超単体の上での積分を計算する。

$$\int_{\{\theta | \sum_j \theta_j = 1, \theta_j \geq 0\}} \prod_j \theta_j^{\alpha_j - 1} d\theta \tag{2}$$

$$= \int_0^1 \theta_1^{\alpha_1 - 1} d\theta_1 \int_0^{1-\theta_1} \theta_2^{\alpha_2 - 1} d\theta_2 \cdots \int_0^{1-\theta_1-\cdots-\theta_{m-2}} \theta_{m-1}^{\alpha_{m-1}-1} d\theta_{m-1} \int_{1-\theta_1-\cdots-\theta_{m-1}}^{1-\theta_1-\cdots-\theta_{m-1}} \theta_m^{\alpha_m - 1} d\theta_m \tag{3}$$

$$= \int_0^1 \theta_1^{\alpha_1 - 1} d\theta_1 \int_0^{1-\theta_1} \theta_2^{\alpha_2 - 1} d\theta_2 \cdots \int_0^{1-\theta_1-\cdots-\theta_{m-2}} \theta_{m-1}^{\alpha_{m-1}-1} (1 - \theta_1 - \cdots - \theta_{m-1})^{\alpha_m - 1} d\theta_{m-1} \tag{4}$$

ここで、$\theta_{m-1} = (1 - \theta_1 - \cdots - \theta_{m-2}) h_{m-1}$ と変数変換する。

$$\int_{\{\theta | \sum_j \theta_j = 1, \theta_j \geq 0\}} \prod_j \theta_j^{\alpha_j - 1} d\theta \tag{5}$$

$$\begin{aligned} &= \int_0^1 \theta_1^{\alpha_1 - 1} d\theta_1 \int_0^{1-\theta_1} \theta_2^{\alpha_2 - 1} d\theta_2 \\ &\cdots \int_0^1 (1 - \theta_1 - \cdots - \theta_{m-2})^{\alpha_{m-1}-1} h_{m-1}^{\alpha_{m-1}-1} \{(1 - \theta_1 - \cdots - \theta_{m-2})(1 - h_{m-1})\}^{\alpha_m - 1} (1 - \theta_1 - \cdots - \theta_{m-2}) dh_{m-1} \end{aligned} \tag{6}$$

$$\begin{aligned} &= \int_0^1 \theta_1^{\alpha_1 - 1} d\theta_1 \int_0^{1-\theta_1} \theta_2^{\alpha_2 - 1} d\theta_2 \\ &\cdots \int_0^{1-\theta_1-\cdots-\theta_{m-3}} \theta_{m-2}^{\alpha_{m-2}-1} (1 - \theta_1 - \cdots - \theta_{m-2})^{\alpha_{m-1}+\alpha_m-1} d\theta_{m-2} \cdot \int_0^1 h_{m-1}^{\alpha_{m-1}-1} (1 - h_{m-1})^{\alpha_m - 1} dh_{m-1} \end{aligned} \tag{7}$$

$$\begin{aligned} &= \int_0^1 \theta_1^{\alpha_1 - 1} d\theta_1 \int_0^{1-\theta_1} \theta_2^{\alpha_2 - 1} d\theta_2 \\ &\cdots \int_0^{1-\theta_1-\cdots-\theta_{m-3}} \theta_{m-2}^{\alpha_{m-2}-1} (1 - \theta_1 - \cdots - \theta_{m-2})^{\alpha_{m-1}+\alpha_m-1} d\theta_{m-2} \cdot B(\alpha_{m-1}, \alpha_m) \end{aligned} \tag{8}$$

式中の B はベータ関数である。次に、$\theta_{m-2} = (1-\theta_1-\cdots-\theta_{m-3})h_{m-2}$ と変数変換する。

$$\int_{\{\theta|\sum_j \theta_j=1,\theta_j\geq 0\}} \prod_j \theta_j^{\alpha_j-1} d\theta \tag{9}$$

$$= \int_0^1 \theta_1^{\alpha_1-1} d\theta_1 \int_0^{1-\theta_1} \theta_2^{\alpha_2-1} d\theta_2$$
$$\cdots \int_0^1 \{(1-\theta_1-\cdots-\theta_{m-3})h_{m-2}\}^{\alpha_{m-2}-1} \{(1-\theta_1-\cdots-\theta_{m-3})(1-h_{m-2})\}^{\alpha_{m-1}+\alpha_m-1} (1-\theta_1-\cdots-\theta_{m-3}) dh_{m-2}$$
$$\cdot B(\alpha_{m-1}, \alpha_m) \tag{10}$$

$$= \int_0^1 \theta_1^{\alpha_1-1} d\theta_1 \int_0^{1-\theta_1} \theta_2^{\alpha_2-1} d\theta_2$$
$$\cdots \int_0^{1-\theta_1-\cdots-\theta_{m-4}} \theta_{m-3}^{\alpha_{m-3}-1} (1-\theta_1-\cdots-\theta_{m-3})^{\alpha_{m-2}+\alpha_{m-1}+\alpha_m-1} d\theta_{m-3}$$
$$\cdot \int_0^1 h_{m-2}^{\alpha_{m-2}-1} (1-h_{m-2})^{\alpha_{m-1}+\alpha_m-1} dh_{m-2} \cdot B(\alpha_{m-1}, \alpha_m) \tag{11}$$

$$= \int_0^1 \theta_1^{\alpha_1-1} d\theta_1 \int_0^{1-\theta_1} \theta_2^{\alpha_2-1} d\theta_2$$
$$\cdots \int_0^{1-\theta_1-\cdots-\theta_{m-4}} \theta_{m-3}^{\alpha_{m-3}-1} (1-\theta_1-\cdots-\theta_{m-3})^{\alpha_{m-2}+\alpha_{m-1}+\alpha_m-1} d\theta_{m-3} \cdot B(\alpha_{m-2}, \alpha_{m-1}+\alpha_m) \cdot B(\alpha_{m-1}, \alpha_m) \tag{12}$$

同様の変数変換を繰り返すと

$$\int_{\{\theta|\sum_j \theta_j=1,\theta_j\geq 0\}} \prod_j \theta_j^{\alpha_j-1} d\theta = B(\alpha_1, \alpha_2+\cdots+\alpha_m) \cdots B(\alpha_{m-2}, \alpha_{m-1}+\alpha_m) B(\alpha_{m-1}, \alpha_m) \tag{13}$$

を得る。式 (13) のベータ関数をガンマ関数によって書き直すと

$$\begin{aligned}\int_{\{\theta|\sum_j \theta_j=1,\theta_j\geq 0\}} \prod_j \theta_j^{\alpha_j-1} d\theta &= \frac{\Gamma(\alpha_1)\Gamma(\alpha_2+\cdots+\alpha_m)}{\Gamma(\alpha_1+\alpha_2+\cdots+\alpha_m)} \cdots \frac{\Gamma(\alpha_{m-2})\Gamma(\alpha_{m-1}+\alpha_m)}{\Gamma(\alpha_{m-2}+\alpha_{m-1}+\alpha_m)} \cdot \frac{\Gamma(\alpha_{m-1})\Gamma(\alpha_m)}{\Gamma(\alpha_{m-1}+\alpha_m)} \\ &= \frac{\prod_j \Gamma(\alpha_j)}{\Gamma(\sum_j \alpha_j)} \end{aligned} \tag{14}$$

となり、この式 (14) の逆数を正規化項とすればよい。