Pattern Recognition and Machine Learning
2. Probability Distribution

2.2 multinomial variables(P74-78)

北海道大学大学院　情報科学研究科 コンピュータサイエンス専攻 情報知識ネットワーク研究室　鈴木康広

イントロダクション

- ディリクレ分布
  - 多値変数とは？
  - 条件付き確率分布
  - ベルヌーイ分布との比較
  - K次元の尤度関数
  - ラグランジュの未定乗数法とその例
  - 1対K法での最尤推定量
  - 多項分布
  - ディリクレ分布の定義
  - ディリクレ分布での平均と分散
  - ディリクレ分布の事後確率

多値変数とは？

・ 二値変数は2つの値のうちどちらかを選ぶ量をを求めるのに利用できる。

・多値変数はK個の値のうち1つを選ぶ場合を考える。

・ここでは特に”1対K法”について考える。

$$x = (0, 0, 1, 0, 0, 0)^T$$

上記は1対K法の例である。K個の要素のうち（例の場合、k=6）、k(k ≦ K) 番目の要素（例の場合、k=3）が1であり、他の要素が0である場合である。

なお、xのk番目の要素が1であることをx_k=1と示す。

条件付き確率分布

xのk番目の要素が1であることをx_kと示すことから$\sum_{k=1}^{K} x_k = 1$ を満たすベクトルである。

この時、x_k=1となる確率をμ_kとすると、

$$p(x \mid \mu) = \prod_{k=1}^{K} \mu_k^{x_k}$$

ただし$\mu=(\mu_1,\mu_2,\dots,\mu_K)^T$ であり、パラメータμ_kは$\mu_k \geqq 0$である事を満たし、このパラメータは確率であるので、$\sum_k \mu_k = 1$ である。

これはベルヌーイ分布を2種類以上に、つまりK種類以上に拡張したものである。

次に、$\sum_x p(x \mid \mu) = \sum_{k=1}^{K} \mu_k = 1$ と $E[x \mid \mu] = \sum_x p(x \mid \mu) x = (\mu_1, \dots, \mu_N)^T = \mu$

の説明を行う。

75ページの式(2.25)から式(2.27)の前まで

ベルヌーイ分布との比較

Xの値	1	0	合計
P(x)	μ	$1-\mu$	1
平均(xp(x))	μ	0	μ
分散($(x-\mu)^2p(x)$)	$(1-\mu)^2\mu$	$(0-\mu)^2$	$\mu(1-\mu)$

$\mu=\{\mu_1,\ldots,\mu_{k-1},\mu_{k+1}\ldots,\mu_K\}$

$\ldots,\mu_K\}$

Xの値	$1(x_k=1)$		合計
P(x)	μ_k	μの要素	1
平均(xp(x))	μ_k	0	$\mu=\{\mu_1,\mu_2,\ldots,\mu_K\}$
分散($(x-\mu)^2p(x)$)	$(1-\mu_k)^2\mu$	$(0-\mu_k)^2(1-\mu_k)$	$\mu(1-\mu)$

$\mu=\{\mu_1,\ldots,\mu_{k-1},\mu_{k+1}\ldots,\mu_K\}$

$\mu=\{\mu_2,\ldots,\mu_K\}$

研究室ゼミ発表(ビショップ)・ディリクレ分布

式(2.27)と式(2.28)の説明

K次元の尤度関数

$$p(D\mid\mu)=\prod_{n=1}^{N}\prod_{k=1}^{K}\mu_k^{x_{nk}}=\prod_{k=1}^{K}\mu_k^{\left(\sum_n x_{nk}\right)}=\prod_{k=1}^{K}\mu_k^{m_k}$$

N個の独立したデータ$x_1,\ldots,x_N$からなるデータ集合Dについて考える。この時の尤度関数は上記のようになる。見ての通りK個の要素を通してN個のデータにのみ依存している。

$$m_k=\sum_n x_{nk}$$

このm$_k$はx$_k$=1である数を表しており、分布に置ける十分統計量(sufficient statistics) と呼ばれる。

式(2.28)の下から式(2.30)の下の段落まで

ラグランジュ未定乗数法

ラグランジュ未定乗数法は複数の変数に1つ以上の制約を与えることで静止点(stationary point、極大点、極小点など)を見つけることである。

一般的に、n個の変数で決まる関数$f(x_1,\ldots,x_n)$が存在する時、各xについて、$g(x_1,\ldots,x_n)=0$の条件が与えられているとする。これは右記の連立方程式を解く事に該当する。

$$\frac{\partial}{\partial x}(f+\lambda g)=0$$

$$\frac{\partial}{\partial y}(f+\lambda g)=0$$

$$\frac{\partial}{\partial z}(f+\lambda g)=0$$

g=0 $\boxed{\frac{\partial}{\partial \lambda}(f+\lambda g)=0}$

次にラグランジュの未定乗数法の例を挙げる。

ラグランジュ未定乗数法の例(1)

$$f(x,y)=x^2+y^2+z^2 \quad g(x,y,z)=2x-y+z+6=0$$

$$\frac{\partial}{\partial x}(f+\lambda g)=0$$

$$\frac{\partial}{\partial y}(f+\lambda g)=0$$

$$\frac{\partial}{\partial z}(f+\lambda g)=0$$

$$\frac{\partial}{\partial \lambda}(f+\lambda g)=0$$

$$\frac{\partial}{\partial x}(x^2+y^2+z^2+\lambda(2x-y+z+6))=0$$

$$2x+2\lambda=0$$

$$x=-\lambda$$

$$\frac{\partial}{\partial y}(x^2+y^2+z^2+\lambda(2x-y+z+6))=0$$

$$2y-\lambda=0$$

$$y=\frac{\lambda}{2}$$

$$\frac{\partial}{\partial z}(x^2+y^2+z^2+\lambda(2x-y+z+6))=0$$

$$2z+\lambda=0$$

$$z=-\frac{\lambda}{2}$$

$$2x-y+z+6=0$$

$$\lambda=2$$

ラグランジュ未定乗数法の例(2)

$$\lambda = 2$$

$$\frac{\partial}{\partial x}(x^2 + y^2 + z^2 + \lambda(2x - y + z + 6)) = 0$$

$$2x + 2\lambda = 0$$

$$\boxed{x = -\lambda}$$

$$\frac{\partial}{\partial y}(x^2 + y^2 + z^2 + \lambda(2x - y + z + 6)) = 0$$

$$2y - \lambda = 0$$

$$\boxed{y = \frac{\lambda}{2}}$$

$$\frac{\partial}{\partial z}(x^2 + y^2 + z^2 + \lambda(2x - y + z + 6)) = 0$$

$$2z + \lambda = 0$$

$$\boxed{z = -\frac{\lambda}{2}}$$

$$(x, y, z) = (-2, 1, -1)$$

1対K法での最尤推定量

ラグランジュの未定乗数法を用いて、P(D|μ)の対数尤度P(D|μ)を最大値を求める。

$$p(D|\mu) = \prod_{n=1}^{N}\prod_{k=1}^{K}\mu_k^{x_{nk}} = \prod_{k=1}^{K}\mu_k^{\left(\sum_n x_{nk}\right)} = \prod_{k=1}^{K}\mu_k^{m_k}$$

$$\ln(P(D|\mu)) = \sum_{k=1}^{K} m_k \ln \mu_k$$

$$\ln(P(D|\mu)) = \sum_{k=1}^{K} m_k \ln \mu_k + \lambda\left(\sum_{k=1}^{K}\mu_k - 1\right)$$

制約条件 $\sum_k \mu_k = 1$

$$\frac{m_k}{\mu_k} + \lambda = 0\,,\ \mu_k = \frac{-m_k}{\lambda}$$

$\lambda = -N$

$$\mu = \frac{m_k}{N}$$

最尤推定量

これはN個の観測値について、x_k=1である数の割合を示している。

式(2.31)の前段落から式(2.32)まで

多項分布

次に、パラメータμと観測値の総数を与えた時の$m_1, m_2, \ldots, m_K$の同時確率を考える。

$$Mult\,(m_1, m_2, \ldots, m_K \mid \mu, N) = \binom{N}{m_1 m_2, \ldots m_K} \prod_{k=1}^{K} \mu_k^{m_k}$$

$$\binom{N}{m_1 m_2, \ldots m_K} = \frac{N!}{m_1! m_2! \ldots m_K!}$$

$$\sum_{k=1}^{K} m_k = N$$

N個の要素から各パラメータを重複を認めた時の選ぶ組み合わせの数。

ディリクレ分布の定義(1)

パラメータ{μ_k}について事前分布の族(family)について考える。

$$Mult\,(m_1, m_2, \ldots, m_K \mid \mu, N) = \binom{N}{m_1 m_2, \ldots m_K} \prod_{k=1}^{K} \mu_k^{m_k}$$

から、その共役分布は

$$p\,(\mu \mid \alpha) \propto \prod_{k=1}^{K} \mu_k^{\alpha_{k-1}}$$

ただしμ_kは$0 \le \mu_k \le 1$であり、$\sum_k \mu_k = 1$であり、αは$\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_K)^T$で表すパラメータである。

ディリクレ分布の定義(2)

76ページの式(2.37)と式(2.38)と式(2.39)について

$$Beta(\mu \mid a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mu^{a-1}(1-\mu)^{b-1}$$

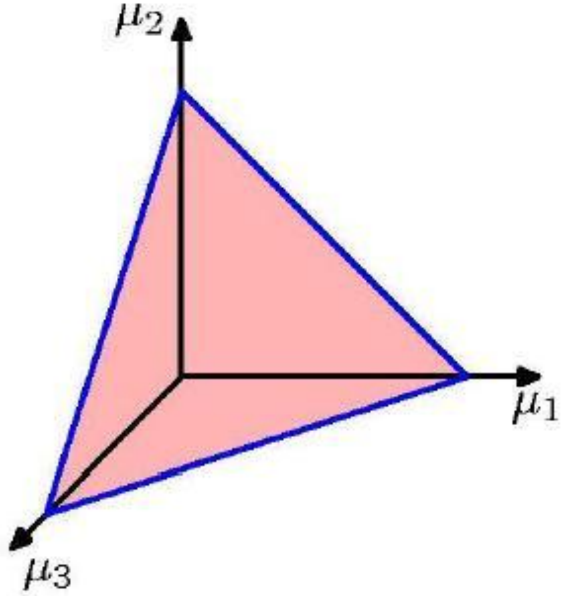
$$Dir(\mu \mid \alpha) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k - 1}$$

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx \qquad \Gamma(a) = (a-1)! \quad \Gamma(1) = 1$$

$$\begin{aligned}
\Gamma(a) &= \int_0^\infty x^{a-1} e^{-x} dx \\
&= \int_0^\infty x^{a-1} (-e^{-x})' dx \\
&= \left[-x^{a-1} e^{-x} \right]_0^\infty + (a-1)\int_0^\infty x^{a-2} e^{-x} dx \\
&= 0 + (a-1)\int_0^\infty x^{a-2} e^{-x} dx \\
&= (a-1)\Gamma(a-1)
\end{aligned}$$

$$\prod_{k=1}^{K} \mu_k^{\alpha_k - 1} = \int_0^1 \mu_1^{\alpha_1 - 1} d\mu_1 \int_0^{1-\mu_1} \mu_2^{\alpha_2 - 1} d\mu_2 \int_0^{1-\mu_1-\mu_2} \mu_3^{\alpha_3 - 1} d\mu_3 \cdots \int_{1-\mu_1\cdots-\mu_{k-1}}^{1-\mu_1\cdots-\mu_{k-1}} \mu_k^{\alpha_k - 1} d\mu_k$$

変数と次元の関係



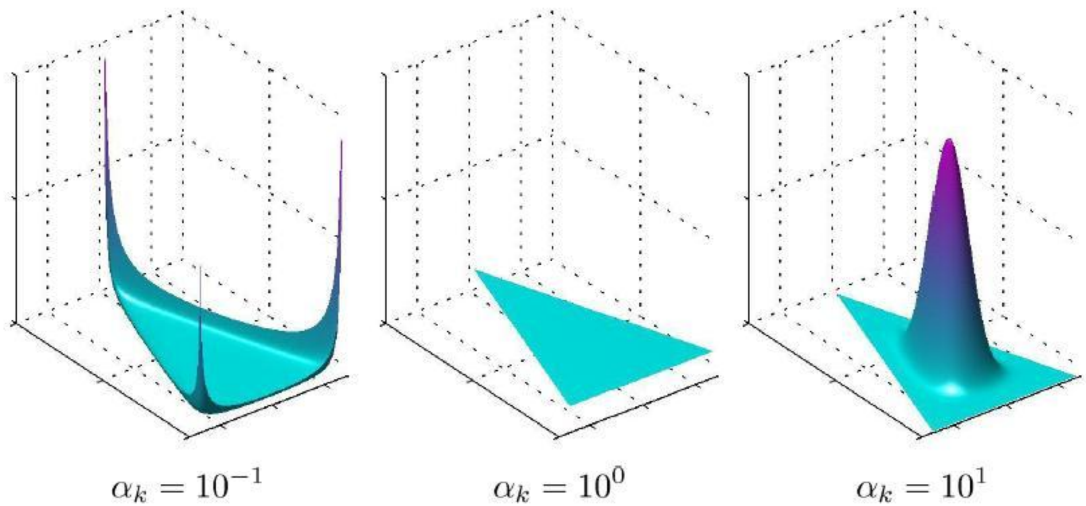


図2.4と図2.5について

左図は3変数のディリクレ分布の有限線形多様体である。$\sum_k \mu_k = 1$からこの単体の範囲を超えることはない。K次元のベクトルである時、K次元の単体(simplex)内に制限される。

右図は3変数のディリクレ分布の一例である。水平軸は平面の座標であり、垂直軸は密度を表している。この時の$a_k=10^{-1}$、$a_k=10^{0}$、$a_k=10^{1}$を図示している。

ディリクレ分布の正規化項(1)

$$\int_{\left\{\mu \mid \sum_k^K \mu_k = 1, \mu_k \ge 0\right\}} \prod_{k=1}^K \mu_k^{\alpha_k - 1}$$

$$= \int_0^1 \mu_1^{\alpha_1-1} d\mu_1 \int_0^{1-\mu_1} \mu_2^{\alpha_2-1} d\mu_2 \int_0^{1-\mu_1-\mu_2} \mu_3^{\alpha_3-1} d\mu_3 \cdots \int_{1-\mu_1\cdots-\mu_{K-1}}^{1-\mu_1\cdots-\mu_{K-1}} \mu_K^{\alpha_K-1} d\mu_K$$

$$= \int_0^1 \mu_1^{\alpha_1-1} d\mu_1 \int_0^{1-\mu_1} \mu_2^{\alpha_2-1} d\mu_2 \cdots \int_0^{1-\mu_1\cdots-\mu_{K-2}} \mu_{K-1}^{\alpha_{K-1}-1} (1-\mu_1-\cdots-\mu_{K-1})^{\alpha_K-1} d\mu_{K-1}$$

$\mu_{K-1} = (1-\mu_1-\cdots-\mu_{K-2})h_{K-1}$と置き換える。

$$\int_{\left\{\mu \mid \sum_k^K \mu_k = 1, \mu_k \ge 0\right\}} \prod_{k=1}^K \mu_k^{\alpha_k - 1}$$

$$= \int_0^1 \mu_1^{\alpha_1-1} d\mu_1 \int_0^{1-\mu_1} \mu_2^{\alpha_2-1} d\mu_2$$

$$\cdots \int_0^1 (1-\mu_1-\cdots\mu_{K-2})^{\alpha_{K-1}-1} h_{K-1}^{\alpha_{K-1}-1} \{(1-\mu_1-\cdots-\mu_{K-2})(1-h_{K-1})\}^{\alpha_K-1} (1-\mu_1-\cdots-\mu_{K-2}) dh_{K-1}$$

ディリクレ分布の正規化項(2)

$$\int_0^1 \mu_1^{\alpha_1-1} d\mu_1 \int_0^{1-\mu_1} \mu_2^{\alpha_2-1} d\mu_2$$

$$\cdots \int_0^1 (1-\mu_1-\cdots\mu_{K-2})^{\alpha_{K-1}-1} h_{K-1}^{\alpha_{K-1}-1} \{(1-\mu_1-\cdots-\mu_{K-2})(1-h_{K-1})\}^{\alpha_K-1}(1-\mu_1-\cdots-\mu_{K-2})dh_{K-1}$$

$$= \int_0^1 \mu_1^{\alpha_1-1} d\mu_1 \int_0^{1-\mu_1} \mu_2^{\alpha_2-1} d\mu_2$$

$$\cdots \int_0^{1-\mu_1-\cdots\mu_{K-3}} \mu_{K-2}^{\alpha_{K-2}-1}(1-\mu_1-\cdots-\mu_{K-2})^{\alpha_{K-1}+\alpha_K-1} d\mu_{K-2} \boxed{\int_0^1 h_{K-1}^{\alpha_{K-1}-1}(1-h_{K-1})^{\alpha_K-1} dh_{K-1}}$$

ベータ関数

$$\int_0^1 Beta(\mu \mid a,b) d\mu = \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mu^{a-1}(1-\mu)^{b-1} d\mu = 1$$

$$\int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \times \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} d\mu = 1$$

$$\boxed{\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 \mu^{a-1}(1-\mu)^{b-1} d\mu}$$

ベータ分布

ディリクレ分布の正規化項(3)

$$\int_0^1 \mu_1^{\alpha_1-1} d\mu_1 \int_0^{1-\mu_1} \mu_2^{\alpha_2-1} d\mu_2 \cdots \int_0^{1-\mu_1-\cdots\mu_{K-3}} \mu_{K-2}^{\alpha_{K-2}-1}(1-\mu_1-\cdots-\mu_{K-2})^{\alpha_{K-1}+\alpha_K-1} d\mu_{K-2} \frac{\Gamma(\alpha_{K-1})\Gamma(\alpha_K)}{\Gamma(\alpha_K+\alpha_{K-1})}$$

$\mu_{K-2}=(1-\mu_1-\cdots-\mu_{K-3})h_{K-2}$ と置き換える。

$$\int_0^1 \mu_1^{\alpha_1-1} d\mu_1 \int_0^{1-\mu_1} \mu_2^{\alpha_2-1} d\mu_2 \cdots \int_0^1 (1-\mu_1-\cdots\mu_{K-3})^{\alpha_{K-2}-1} h_{K-2}^{\alpha_{K-2}-1}\{(1-\mu_1-\cdots-\mu_{K-3})(1-h_{K-2})\}^{\alpha_{K-1}+\alpha_K-1}(1-\mu_1-\cdots-\mu_{K-3})dh_{K-2} \times \frac{\Gamma(\alpha_{K-1})\Gamma(\alpha_K)}{\Gamma(\alpha_K+\alpha_{K-1})}$$

$$= \int_0^1 \mu_1^{\alpha_1-1} d\mu_1 \int_0^{1-\mu_1} \mu_2^{\alpha_2-1} d\mu_2 \cdots \int_0^{1-\mu_1-\cdots\mu_{K-4}} \mu_{K-3}^{\alpha_{K-3}-1}(1-\mu_1-\cdots-\mu_{K-3})^{\alpha_{K-2}+\alpha_{K-1}+\alpha_K-1} d\mu_{K-3} \underbrace{\boxed{\int_0^1 h_{K-2}^{\alpha_{K-2}-1}(1-h_{K-2})^{\alpha_{K-1}+\alpha_K-1} dh_{K-2}}}_{\text{ベータ関数}} \times \underbrace{\boxed{\frac{\Gamma(\alpha_{K-1})\Gamma(\alpha_K)}{\Gamma(\alpha_{K-1}+\alpha_K)}}}_{\text{ベータ関数}}$$

ディリクレ分布の正規化項(4)

$$\int_{\left\{\mu \middle| \sum_k^K \mu_k = 1, \mu_k \geq 0\right\}} \prod_{k=1}^{K} \mu_k^{\alpha_k - 1} d\mu_k = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2 + \cdots + \alpha_K)}{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_K)} \times \cdots \frac{\Gamma(\alpha_{K-2})\Gamma(\alpha_{K-1} + \alpha_K)}{\Gamma(\alpha_{K-2} + \alpha_{K-1} + \alpha_K)} \times \frac{\Gamma(\alpha_K)\Gamma(\alpha_{K-1})}{\Gamma(\alpha_K + \alpha_{K-1})}$$

$$= \frac{\prod_{k=1}^{K} \Gamma(\alpha_k)}{\Gamma\left(\sum_{k=1}^{K} \alpha_k\right)}$$

となるので、ディリクレ分布は以下のようになる。

$$Dir(\mu \mid \alpha) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k - 1}$$

ディリクレ分布の正規化項(4)

$$\int_{\left\{\mu \mid \sum_{k}^{K}\mu_k=1,\mu_k\geq 0\right\}} \prod_{k=1}^{K}\mu_k^{\alpha_k-1}d\mu_k = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2+\cdots+\alpha_K)}{\Gamma(\alpha_1+\alpha_2+\cdots+\alpha_K)}\times\cdots\frac{\Gamma(\alpha_{K-2})\Gamma(\alpha_{K-1}+\alpha_K)}{\Gamma(\alpha_{K-2}+\alpha_{K-1}+\alpha_K)}\times\frac{\Gamma(\alpha_K)\Gamma(\alpha_{K-1})}{\Gamma(\alpha_K+\alpha_{K-1})}$$

$$=\boxed{\frac{\prod_{k=1}^{K}\Gamma(\alpha_k)}{\Gamma\left(\sum_{k=1}^{K}\alpha_k\right)}}$$

となるので、ディリクレ分布は以下のようになる。

$$Dir(\mu\mid\alpha)=\frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_K)}\boxed{\prod_{k=1}^{K}\mu_k^{\alpha_k-1}}$$

ディリクレ分布の平均

$$\int_{\left\{\mu \mid \sum_k^K \mu_k = 1, \mu_k \ge 0\right\}} \frac{\Gamma(\alpha_1+\alpha_2+\cdots+\alpha_K)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k-1} d\mu_k = 1$$

$$E[\mu_x] = \int_{\left\{\mu \mid \sum_k^K \mu_k = 1, \mu_k \ge 0\right\}} \frac{\Gamma(\alpha_1+\alpha_2+\cdots+\alpha_K)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k-1} \times \mu_x d\mu_k$$

$$= \int_{\left\{\mu \mid \sum_k^K \mu_k = 1, \mu_k \ge 0\right\}} \frac{\Gamma(\alpha_1+\alpha_2+\cdots+\alpha_K)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_K)} \left(\prod_{k=1, k\ne x}^{K} \mu_k^{\alpha_k-1} d\mu_k \right) \times \left(\mu_x^{\alpha_x} d\mu_x \right)$$

$$= \int_{\left\{\mu \mid \sum_k^K \mu_k = 1, \mu_k \ge 0\right\}} \frac{\dfrac{\Gamma(\alpha_1+\alpha_2+\cdots\alpha_x+1+\cdots+\alpha_K)}{\alpha_0}}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots+\dfrac{\Gamma(\alpha_x+1)}{\alpha_k}+\Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k-1} d\mu_k \times \left(\mu_x^{\alpha_x} d\mu_x \right)$$

$$= \frac{\alpha_x}{\alpha_0} \int_{\left\{\mu \mid \sum_k^K \mu_k = 1, \mu_k \ge 0\right\}} \frac{\Gamma(\alpha_1+\cdots+\alpha_x+1+\cdots\alpha_K)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots+\Gamma(\alpha_x+1)+\Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k-1} d\mu_k \times \left(\mu_x^{\alpha_x} d\mu_x \right)$$

$$= \frac{\alpha_x}{\alpha_0}$$

ディリクレ分布の平均

$$\boxed{\int_{\left\{\mu \middle| \sum_k^K \mu_k=1,\mu_k\ge 0\right\}} \frac{\Gamma(\alpha_1+\alpha_2+\cdots+\alpha_K)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k-1} d\mu_k = 1}$$

$$E[\mu_x] = \int_{\left\{\mu \middle| \sum_k^K \mu_k=1,\mu_k\ge 0\right\}} \frac{\Gamma(\alpha_1+\alpha_2+\cdots+\alpha_K)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k-1} \times \mu_x d\mu_k$$

$$= \int_{\left\{\mu \middle| \sum_k^K \mu_k=1,\mu_k\ge 0\right\}} \frac{\Gamma(\alpha_1+\alpha_2+\cdots+\alpha_K)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_K)} \left(\prod_{k=1,k\neq x}^{K} \mu_k^{\alpha_k-1} d\mu_k\right) \times \left(\mu_x^{\alpha_x} d\mu_x\right)$$

$$= \int_{\left\{\mu \middle| \sum_k^K \mu_k=1,\mu_k\ge 0\right\}} \frac{\dfrac{\Gamma(\alpha_1+\alpha_2+\cdots\alpha_x+1+\cdots+\alpha_K)}{\alpha_0}}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots+\dfrac{\Gamma(\alpha_x+1)}{\alpha_k}+\Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k-1} d\mu_k$$

= 1

$$= \frac{\alpha_x}{\alpha_0} \boxed{\int_{\left\{\mu \middle| \sum_k^K \mu_k=1,\mu_k\ge 0\right\}} \frac{\Gamma(\alpha_1+\cdots+\alpha_x+1+\cdots\alpha_K)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots+\Gamma(\alpha_x+1)+\Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k-1} d\mu_k \times \left(\mu_x^{\alpha_x} d\mu_x\right)}$$

$$= \frac{\alpha_x}{\alpha_0}$$

ディリクレ分布の分散

$$E[\mu_x^2] = \int_{\left\{\mu | \sum_k^K \mu_k = 1, \mu_k \geq 0\right\}} \frac{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_K)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k - 1} \times \mu_x^2 d\mu_k$$

$$= \int_{\left\{\mu | \sum_k^K \mu_k = 1, \mu_k \geq 0\right\}} \frac{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_K)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_K)} \left(\prod_{k=1, k \neq x}^{K} \mu_k^{\alpha_k - 1} d\mu_k \right) \times \left(\mu_x^{\alpha_x + 1} d\mu_x \right)$$

$$= \int_{\left\{\mu | \sum_k^K \mu_k = 1, \mu_k \geq 0\right\}} \frac{\dfrac{\Gamma(\alpha_1 + \alpha_2 + \cdots \alpha_x + 2 + \cdots + \alpha_K)}{\alpha_0(\alpha_0 + 1)}}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots + \dfrac{\Gamma(\alpha_x + 2)}{\alpha_x(\alpha_x + 1)} + \Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k - 1} d\mu_k \times \left(\mu_x^{\alpha_x + 1} d\mu_x \right)$$

$$= \frac{\alpha_x(\alpha_x + 1)}{\alpha_0(\alpha_0 + 1)} \int_{\left\{\mu | \sum_k^K \mu_k = 1, \mu_k \geq 0\right\}} \frac{\Gamma(\alpha_1 + \cdots + \alpha_x + 2 + \cdots \alpha_K)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots + \Gamma(\alpha_x + 2) + \Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k - 1} d\mu_k \times \left(\mu_x^{\alpha_x} d\mu_x \right)$$

$$\boxed{\mathrm{var}(\mu_x) = \frac{\alpha_x(\alpha_x + 1)}{\alpha_0(\alpha_0 + 1)} - \left(\frac{\alpha_x}{\alpha_0} \right)^2 = \frac{\alpha_x \alpha_0 (\alpha_x + 1) - \alpha_x^2 (\alpha_0 + 1)}{\alpha_0^2 (\alpha_0 + 1)} = \frac{\alpha_x(\alpha_0 - \alpha_x)}{\alpha_0^2(\alpha_0 + 1)}}$$

ディリクレ分布の分散

$$E[\mu_x^2] = \int_{\left\{\mu \mid \sum_k^K \mu_k = 1, \mu_k \geq 0\right\}} \frac{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_K)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k - 1} \times \mu_x^2 d\mu_k$$

$$= \int_{\left\{\mu \mid \sum_k^K \mu_k = 1, \mu_k \geq 0\right\}} \frac{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_K)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_K)} \left(\prod_{k=1, k \neq x}^{K} \mu_k^{\alpha_k - 1} d\mu_k \right) \times \left(\mu_x^{\alpha_x + 1} d\mu_x \right)$$

$$= \int_{\left\{\mu \mid \sum_k^K \mu_k = 1, \mu_k \geq 0\right\}} \frac{\dfrac{\Gamma(\alpha_1 + \alpha_2 + \cdots \alpha_x + 2 + \cdots + \alpha_K)}{\alpha_0(\alpha_0 + 1)}}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots + \dfrac{\Gamma(\alpha_x + 2)}{\alpha_x(\alpha_x + 1)} + \Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k - 1} \quad \mu_x)$$

= 1

$$= \frac{\alpha_x(\alpha_x + 1)}{\alpha_0(\alpha_0 + 1)} \boxed{\int_{\left\{\mu \mid \sum_k^K \mu_k = 1, \mu_k \geq 0\right\}} \frac{\Gamma(\alpha_1 + \cdots + \alpha_x + 2 + \cdots \alpha_K)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots + \Gamma(\alpha_x + 2) + \Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k - 1} d\mu_k \times \left(\mu_x^{\alpha_x} d\mu_x \right)}$$

$$\mathrm{var}(\mu_x) = \frac{\alpha_x(\alpha_x + 1)}{\alpha_0(\alpha_0 + 1)} - \left(\frac{\alpha_x}{\alpha_0} \right)^2 = \frac{\alpha_x \alpha_0 (\alpha_x + 1) - \alpha_x^2 (\alpha_0 + 1)}{\alpha_0^2 (\alpha_0 + 1)} = \frac{\alpha_x(\alpha_0 - \alpha_x)}{\alpha_0^2(\alpha_0 + 1)}$$

Last

ディリクレ分布の事後確率

$$Dir(\mu|\alpha) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_K)}\prod_{k=1}^{K}\mu_k^{\alpha_k-1} \quad Mult(m_1, m_2, \ldots, m_K|\mu, N) = \binom{N}{m_1 m_2, \ldots m_K}\prod_{k=1}^{K}\mu_k^{m_k}$$

$$p(\mu|D,\alpha) \propto p(D|\mu)p(\mu|\alpha) \propto \prod_{k=1}^{K}\mu^{a_k+m_k-1}$$

事前分布としてのディリクレ分布と尤度関数としての多項分布の積からパラメータμ_kの分布は共役性を持つ分布になることから、

$$p(\mu|D,\alpha) = Dir(\mu|\alpha+m)$$
$$= \frac{\Gamma(\alpha_0+N)}{\Gamma(\alpha_1+m_1)\cdots\Gamma(\alpha_K+m_K)}\prod_{k=1}^{K}\mu_k^{\alpha_k+m_k-1}$$

となる。この時のα_kはベータ分布の時の有効観測数mと同意である。

研究室ゼミ発表(ビショップ)・ディリクレ分布

77ページの式(2.40)について