2.2 多値変数

前回は、二項分布に対してベータ分布が共役事前分布であることを示した。本節では、多項分布に対してディリクレ分布が共役事前分布になることを示す。

K 種類の値を取りうるデータが一つ得られるときの確率は次の通り。

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

← カテゴリ分布 と呼ばれたりもする。

データが N 個得られ場合は、

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

$$= \prod_{k=1}^{K} \mu_k^{m_k}$$

← カテゴリ k が出た回数。

最尤推定による μ の解は、

$$\mu_k^{ML} = \frac{m_k}{N}$$

となる。（これまで通り尤度最小化すれば良いので略）

ここで、カテゴリが出た回数を確率変数とした時の確率関数が多項分布と呼ばれる。

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

2.2.1 ディリクレ分布

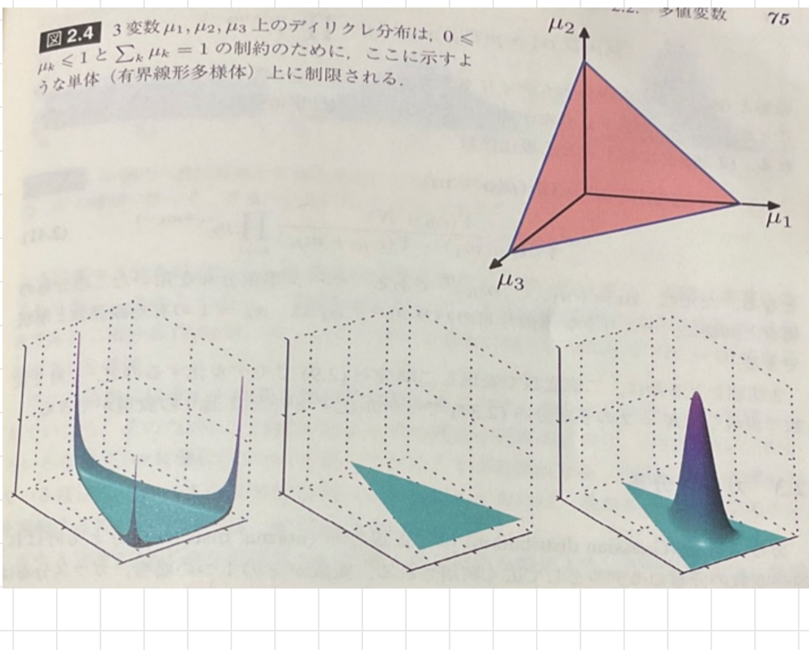
多項分布のパラメータ μ の事前分布として、ディリクレ分布を導入する。天下り的だが、次の性質を持つ分布が、μ の共役事前分布となる。

$$p(\mu|\alpha) \propto \prod_{k=1}^{K} \mu_k^{\alpha_k - 1}$$

(ただし、μ は $0 \leq \mu \leq 1$ かつ $\sum_k \mu_k = 1$ を満たす)

これを正規化すると、次の分布が求まる。~~(※証明)~~ やらない

$$\mathrm{Dir}(\mu|\alpha) = \frac{\Gamma\left(\sum_{k=1}^{K} \alpha_k\right)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k - 1}$$



実際に事後分布がディリクレ分布になることを示す。

$$p(\mu \mid D, \alpha) \propto \underbrace{p(D \mid \mu, \alpha)}_{\text{多項分布}} \underbrace{p(\mu \mid \alpha)}_{\text{ディリクレ分布}}$$

$$= \underbrace{\binom{N}{m_1 \cdots m_K} \prod_{k=1}^{K} \mu_k^{m_k}}_{} \cdot \underbrace{\frac{\Gamma(\sum_{k=1}^{K} \alpha_k)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k - 1}}_{}$$

$$\propto \prod_{k=1}^{K} \mu_k^{\alpha_k + m_k - 1}$$

これを正規化すると、先程の議論同様に、

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

となり、ディリクレ分布となることがわかる。

2.3 ガウス分布

ガウス分布のいくつかの性質について取り上げる。2.3では、多変量ガウス分布が、座標変換により、独立な1変数ガウス分布の積に分解できることを示す。2.3.1では、多変量ガウス分布の条件付き分布がガウス分布となることを示す。2.3.2では、周辺分布もガウス分布となることを示す。

まず、ガウス分布

$$N(x|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}$$

が、正規直交行列 U による変換

$$y = U(x-\mu)$$

により、

$$p(y) = \prod_{j=1}^{D} \underbrace{\frac{1}{(2\pi\lambda_i)^{1/2}} \exp\left\{-\frac{y_j^2}{2\lambda_j}\right\}}_{\text{1変数ガウス分布}}$$

となることを示す。

過去に、分散共分散行列が半正定値であることを示した。
固有値0については後で考察する。

[証]

まず、Σ は正定値の対照行列であるので、
固有値と固有ベクトル λ_i, u_i を用いて、スペクトル分解できる。

$$\Sigma = \sum_{i=1}^{D} \lambda_i u_i u_i'$$

Σ^{-1} についても、同様にスペクトル分解でき、

$$\Sigma^{-1} = \sum_{i=1}^{D} \underbrace{\frac{1}{\lambda_i}}_{\text{固有値は逆数}} \underbrace{u_i u_i'}_{\text{固有ベクトルはそのまま}}$$

これを使うと、ガウス分布の指数部分の二次形式は、
次のように変形できる。

$$\Delta^2 = (x - \mu)' \Sigma^{-1} (x - \mu)$$

$$= \sum_{i=1}^{D} \underbrace{(x - \mu)' u_i}_{y_i} \frac{1}{\lambda_i} \underbrace{u_i' (x - \mu)}_{y_i}$$

$$= \sum_{i=1}^{D} \frac{y_i^2}{\lambda_i}$$

ここで、$y_i = u_i' (x - \mu)$ で変換している。
i をまとめると、

$$y = \underbrace{U}_{\text{正規直交行列}} (x - \mu)$$

これは幾何的に見ると、μ だけ平行移動し、U で
回転した座標に変換したことを示している。

最後に $p(x)$ から $p(y)$ の変換を試みる。

$$p(y) = p(x) \Big|_{y = U(x-\mu)} \cdot |J|$$

$$= \frac{1}{(2\pi)^{D/2}} \underbrace{\frac{1}{|\Sigma|^{1/2}}}_{①} \underbrace{\exp\left\{ -\frac{1}{2} \sum_{j=1}^{D} \frac{y_j^2}{\lambda_j} \right\}}_{②} \underbrace{|J|}_{③}$$

① について

Σ の行列式は固有値の積と一致するので、

$$\frac{1}{|\Sigma|^{\frac{1}{2}}} = \prod_{j=1}^{D} \frac{1}{\lambda_j^{\frac{1}{2}}}$$

② について

Σ を exp の外側に出して、

$$② = \prod_{j=1}^{D} \exp\left\{-\frac{y_j^2}{2\lambda_j}\right\}$$

③ について

$$\mathbf{y} = U(\mathbf{x} - \boldsymbol{\mu}) \text{ より、}$$

$$\begin{aligned} \mathbf{x} &= U^{-1}\mathbf{y} + \boldsymbol{\mu} \\ &= U^{T}\mathbf{y} + \boldsymbol{\mu} \qquad (\because U \text{ の直交性}) \end{aligned}$$

したがって

$$|J| = \left|\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\right| = |U^{T}| = 1 \qquad (\because U \text{ の正規性、詳細は教科書})$$

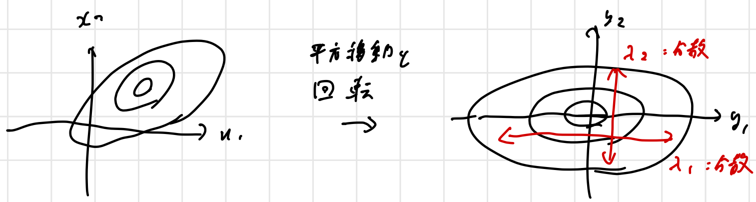
以上を合わせると、

$$p(\mathbf{y}) = \frac{1}{(2\pi)^{1/2D}} \underbrace{\prod_{j=1}^{D}\frac{1}{\lambda_j}}_{①} \underbrace{\prod_{j=1}^{D}\exp\left\{-\frac{y_j^2}{2\lambda_j}\right\}}_{②} \cdot \underbrace{1}_{③}$$

$$= \prod_{j=1}^{D} \frac{1}{(2\pi\lambda_j)^{1/2}} \exp\left\{-\frac{y_j^2}{2\lambda_j}\right\}$$

□

変数変換のイメージ



Σ の固有値 λ_j が 0 になるとき、ガウス分布はある面につぶれていることがわかる。

次に、ガウス分布の1次モーメントと2次モーメントを求めているが、既に学習済みの内容なので詳しくは扱わない。

1次モーメントについて
積分区間内で、zの1次の項は奇関数なので0となる。0次の項はガウス分布の全積分で1が出てくるので、係数μだけ現れる。

2次モーメントについて
zの0次の項：ガウスの全積分により、係数部分$\mu\mu'$が出てくる。
zの1次の項：奇関数で0となる。
zの2次の項：Σが出てくる（下記）

(P81 - 行目の式)

$$= \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \int \exp\left\{-\frac{1}{2}\underbrace{z'\Sigma^{-1}z}_{\Delta^2}\right\} zz' \, dz$$

$y = Uz$ の変換をすると、

$$\Delta^2 = \sum \frac{y_j^2}{\lambda_j} \qquad zz' = (U'y)(y'U) = U'yy'U$$

$$= \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} U' \underbrace{\int \exp\left\{-\frac{1}{2}\sum_j \frac{y_j^2}{\lambda_j}\right\} yy' \, dy}_{K} \, U$$

$$yy' = \begin{pmatrix} y_1y_1 & y_1y_2 & \cdots & y_1y_n \\ \vdots & & & \\ y_ny_1 & & & y_ny_n \end{pmatrix}$$

$(y_1^2\ y_2^2\ \ )I$

Kの部分は行列となっている。s, t成分だけ取り出すと、

(s ≠ t のとき)

$$K_{s,t} = \iint\cdots\int \exp\left\{-\frac{1}{2}\sum_j \frac{y_j^2}{\lambda_j}\right\} y_s y_t \; dy_1 \cdots dy_s \cdots dy_t \cdots dy_D$$

sの積分から実行

$$= \iint\cdots \underbrace{\int \exp\left\{-\frac{1}{2}\sum_j \frac{y_j^2}{\lambda_j}\right\} y_s \, dy_s}_{y_s\text{に関して奇関数}} \; y_t \, dy_1 \cdots dy_t \cdots dy_D$$

$$= 0$$

(s = t のとき)

$$K_{ss} = \iint\cdots\int \exp\left\{-\frac{1}{2}\sum_j \frac{y_j^2}{\lambda_j}\right\} y_s^2 \; dy_1 \cdots dy_D$$

$$= \iint\cdots\int \prod_j^D \exp\left\{-\frac{y_j^2}{2\lambda_j}\right\} y_s^2 \; dy_1 \cdots dy_D$$

$y_1 \sim y_D$で関係する項ごとに積分実行。

$$= \int \exp\left\{-\frac{y_1^2}{2\lambda_1}\right\} dy_1$$

$$+ \cdots$$

$$+ \int \exp\left\{-\frac{y_s^2}{2\lambda_s}\right\} y_s^2 \, dy_s$$

$$+ \cdots$$

$$+ \int \exp\left\{-\frac{y_D^2}{2\lambda_D}\right\} dy_D$$

Kの係数部分を含めると、

$$
\frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} K_{ss} = \int \frac{1}{(2\pi\lambda_1)^{1/2}} \exp\left\{-\frac{y_1^2}{2\lambda_1}\right\} dy_1
$$

$$
+ \cdots
$$

$$
+ \int \frac{1}{(2\pi\lambda_s)^{1/2}} \exp\left\{-\frac{y_s^2}{2\lambda_s}\right\} y_s^2 \, dy_s
$$

$$
+ \cdots
$$

$$
+ \int \frac{1}{(2\pi\lambda_D)^{1/2}} \exp\left\{-\frac{y_D^2}{2\lambda_D}\right\} dy_D
$$

$$
= 0 + \cdots + 0 + \lambda_s + 0 \cdots + 0
$$

$$
= \lambda_s
$$

したがって、

$$
\frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} K = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_D \end{pmatrix}
$$

これより、(P81 一行目)は、

$$
\frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} U' \underbrace{\int \exp\left\{-\frac{1}{2}\sum_j \frac{y_j^2}{\lambda_j}\right\} yy' \, dy}_{= K} U
$$

$$
= U' \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_D \end{pmatrix} U
$$

$$
= \Sigma
$$

Σ は U により次のように対角化される.

$$
U \Sigma U' = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_D \end{pmatrix}
$$

よって、

$$
\Sigma = U' \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_D \end{pmatrix} U
$$

z の 2次式について おわり.

共分散行列のパラメータの個数について
ガウス分布でモデル化するとき、次元が大きいと分散行列のパラメータ数が非常に多くなってしまう。そこで、共分散行列を対角と仮定したり、等方共分散と仮定したりすることで、パラメータ数を減らせる。

一般　　　　対角　　　　等方共分散

パラメータ数：　$\frac{1}{2}D(D+3)$　　　　$2D$　　　　$D+1$

2.3.1 条件付きガウス分布

多変量ガウス分布をいくつかの変量で条件付けた条件付き分布も、ガウス分布に従うことを示す。その際に、精度行列を用いると式がシンプルになることも確かめる。

x を $x = \begin{pmatrix} x_a \\ x_b \end{pmatrix}$ のように分解し、対応する

μ, Σ も次の同様に分解する。

$$\mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$$

Λ を次のように定義し、精度行列と呼ぶ

$$\Lambda = \Sigma^{-1}$$

Λ を分解する. $\quad \Lambda = \begin{pmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{pmatrix}$

$p(x_a \mid x_b)$ を求めるには. $p(x_a, x_b)$ を x_a の分布の形に変形し.正規化すれば良い.

$p(x_a, x_b)$ の指数部分の二次形式を Δ^2 とおくと.

$$\Delta^2 = -\frac{1}{2}(x-\mu)' \Sigma^{-1} (x-\mu)$$

$\boxed{\text{イメージ}}$

$$\begin{pmatrix} a' & b' \end{pmatrix} \begin{pmatrix} aa & ab \\ ba & bb \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

$$= \begin{pmatrix} a'\,aa + b'\,ba & a'\,ab + b'\,bb \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

$$= a'\,aa\,a + b'\,ba\,a + a'\,ab\,b + b'\,bb\,b$$

$$= -\frac{1}{2}(x_a-\mu_a)'\Lambda_{aa}(x_a-\mu_a) - \frac{1}{2}(x_a-\mu_a)'\Lambda_{ab}(x_b-\mu_b)$$
$$-\frac{1}{2}(x_b-\mu_b)'\Lambda_{ba}(x_a-\mu_a) - \frac{1}{2}(x_b-\mu_b)'\Lambda_{bb}(x_b-\mu_b) \qquad (2.70)$$

これを x_a に関する二次形式

$$-\frac{1}{2}(x_a - \mu_{a|b})' \Sigma_{a|b}^{-1} (x_a - \mu_{a|b}) + \text{const.} \qquad (2.71)$$

の形に変形したい.

x_a の2次の項を (2.70) と比較すると.

$$x_a' \Sigma_{a|b}^{-1} x_a = x_a' \Lambda_{aa} x_a$$

より、即座に $\Sigma_{a|b}^{-1} = \Lambda_{aa}$ が分かる。

次に1次の項を比較する。

『(2.71)の1次の項』=『(2.70)の1次の項』より

転置とると同じ

$$x_a' \Sigma_{a|b}^{-1} \mu_{a|b} = \frac{1}{2} x_a' \Lambda_{aa} \mu_a + \frac{1}{2} \mu_a' \Lambda_{aa} x_a$$

$$- \frac{1}{2} x_a' \Lambda_{ab} (x_b - \mu_b)$$

転置とると同じ

$$- \frac{1}{2} (x_b - \mu_b)' \Lambda_{ba} x_a$$

$$= x_a' \Lambda_{aa} \mu_a - x_a' \Lambda_{ab} (x_b - \mu_b)$$

$$= x_a' \{ \Lambda_{aa} \mu_a - \Lambda_{ab} (x_b - \mu_b) \}$$

$$\therefore \quad \Lambda_{aa} \mu_{a|b} = \Lambda_{aa} \mu_a - \Lambda_{ab} (x_b - \mu_b)$$

$$\Leftrightarrow \quad \mu_{a|b} = \Lambda_{aa}^{-1} \{ \Lambda_{aa} \mu_a - \Lambda_{ab} (x_b - \mu_b) \}$$

$$= \mu_a - \Lambda_{aa}^{-1} \Lambda_{ab} (x_b - \mu_b)$$

よって、$p(x_a | x_b)$ は、次の平均 $\mu_{a|b}$ 共分散行列 $\Sigma_{a|b}$ のガウス分布に従う。

$$\begin{cases} \mu_{a|b} = \mu_a - \Lambda_{aa}^{-1} \Lambda_{ab} (x_b - \mu_b) \\ \Sigma_{a|b} = \Lambda_{aa}^{-1} \end{cases}$$

シューア補行列を使った逆行列の公式を使って、

$\Sigma_{a|b}$ を Σ で陽に表すことができるらしいが、

詳しくは扱わない。