第2章　確率分布

2009/06/05(金)
PRMLゼミ
Ｍ１　山田孝太郎

BehaviorInNetworks
都市生活行動学研究室

内容

1. 二値変数の確率分布
   – ベータ分布
2. 多値変数の確率分布
   – ディリクレ分布
3. ガウス分布
   – 条件付きガウス分布
   – 周辺ガウス分布
   – ガウス変数に対するベイズの定理
   – ガウス分布の最尤推定
   – 逐次推定（来週）

はじめに

BehaviorInNetworks
都市生活行動学研究室

- 密度推定
  - 観測値の集合{X_1,…,X_n}が与えられた時に，確率変数Xの従う確率分布を求めること。
  - 候補は無数にあるので，多項式曲線フィッティングやモデル選択問題と関係がある。
- パラメトリックな確率分布
  - パラメータによって定まる確率分布を求める
    - 頻度主義的アプローチ:尤度関数などの基準最適化
    - ベイズ主義的アプローチ:事前分布を導入し，パラメータが得られた時の事後分布を求める。
- ノンパラメトリックな確率分布
  - 分布の形状を制限しない。
  - データ集合の大きさに形状が依存
  - パラメータはあっても分布の複雑さの調整

1. 二値変数

BehaviorInNetworks
都市生活行動学研究室

- **ベルヌーイ分布**: 二値確率変数x∈{0,1}が従う分布

例） 歪なコイン投げ

表(x=1)が出る確率 μ($0 \leq \mu \leq 1$)のコイン
裏(x=0)が出る確率 $1-\mu$
なので，xの値をとる確率は

$$\mathrm{Bern}(x|\mu) = \mu^x(1-\mu)^{1-x}$$

となる。これがベルヌーイ分布の確率密度関数。

- **二項分布**: 表が出る回数mの確率分布

$$\mathrm{Bin}(m|N,\mu) = \binom{N}{m}\mu^m(1-\mu)^{N-m}.$$

ただし.

$$\binom{N}{m} \equiv \frac{N!}{(N-m)!m!}$$

$$\mathbb{E}[m] \equiv \sum_{m=0}^{N} m\mathrm{Bin}(m|N,\mu) = N\mu$$

$$\mathrm{var}[m] \equiv \sum_{m=0}^{N}(m-\mathbb{E}[m])^2\,\mathrm{Bin}(m|N,\mu) = N\mu(1-\mu)$$

というN個からm個取り出す組み合わせの数

1. 二値変数

データ集合 $D = \{x_1, \ldots, x_n\}$ が取れたとき，尤度関数は,

$$p(\mathcal{D}|\mu) = \prod_{n=1}^{N} p(x_n|\mu) = \prod_{n=1}^{N} \mu^{x_n}(1-\mu)^{1-x_n}$$

となる。

ここで，頻度主義的にμを求める･･･尤度関数最大化！

尤度関数の対数をとって,

$$\ln p(\mathcal{D}|\mu) = \sum_{n=1}^{N} \ln p(x_n|\mu) = \sum_{n=1}^{N} \{x_n \ln \mu + (1-x_n)\ln(1-\mu)\}.$$

μで微分し，微分係数を=0とおいてμの最尤推定量を求めると

$$\boxed{\mu_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N} x_n}$$

という，サンプル平均が求められる。

1．1　ベータ分布

•頻度主義的では，データ集合が少ない時，過学習の可能性
（ex.3回投げて3回表が出る）
⇒ベイズ主義的アプローチ：パラメータの事前分布 $p(\mu)$ を設定

•どう設定するか？

•<u>事後分布∝事前分布×尤度関数</u>

⇒事前分布と事後分布は同じ関数形（共役性）

•ベータ分布

μと（1ーμ）のべき乗の形

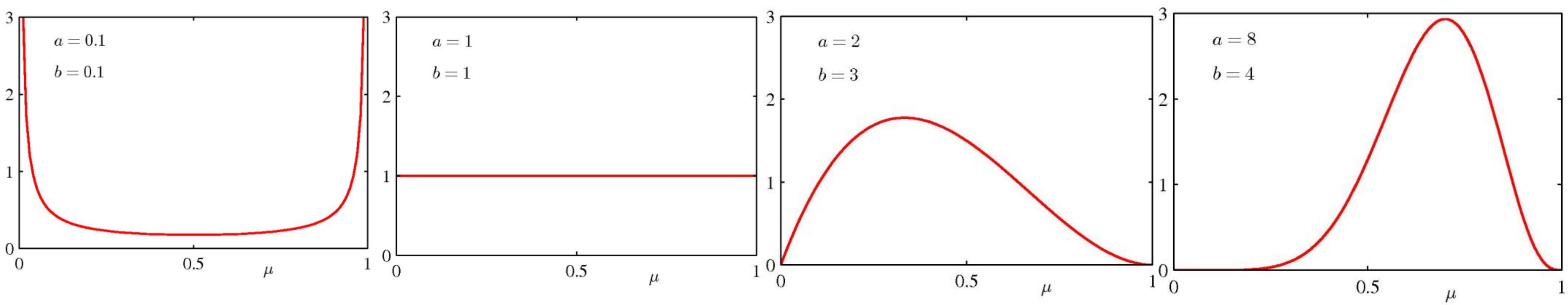
$$\mathrm{Beta}(\mu|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1}.$$

$$\mathbb{E}[\mu] = \frac{a}{a+b}$$

$$\mathrm{var}[\mu] = \frac{ab}{(a+b)^2(a+b+1)}$$

•パラメータa,bはμの分布を決めるので，超パラメータと呼ばれる。

いろいろなa,bのときのベータ分布の形



1．1　ベータ分布

BehaviorInNetworks
都市生活行動学研究室

•事後分布∝事前分布×尤度関数（二項分布）としてμに関する項だけにし，正規化係数をつけると

$$p(\mu|m,l,a,b) = \frac{\Gamma(m+a+l+b)}{\Gamma(m+a)\Gamma(l+b)}\mu^{m+a-1}(1-\mu)^{l+b-1}$$

となり、l=N-m（裏の出た回数）である。

•つまり，あるデータ集合が得られた時，aとbをそれぞれm,lだけ増やせば，事前分布から事後分布が求められる。
•このときa,bは有効観測数として解釈できる。
•そして，得られた事後分布は次の試行の事前分布となりうる。
⇒逐次学習のアプローチ
例）あるデータ集合が得られたときの次の試行で表が出る確率

$$p(x=1|\mathcal{D}) = \int_0^1 p(x=1|\mu)p(\mu|\mathcal{D})\,\mathrm{d}\mu = \int_0^1 \mu p(\mu|\mathcal{D})\,\mathrm{d}\mu = \mathbb{E}[\mu|\mathcal{D}]$$

結局，

$$p(x=1|\mathcal{D}) = \frac{m+a}{m+a+l+b}$$

2. 多値変数

BehaviorInNetworks
都市生活行動学研究室

•異なるK個の可能な状態を取りうる確率変数についての記述

1対K法:
変数は要素の一つx_kが1で残りは0と記述されるようなベクトルになる。たとえばK=6でx_3 =1のときは

$$\mathbf{x} = (0, 0, 1, 0, 0, 0)^{\mathrm{T}}$$

Xの分布は, X_k=1となる確率をμ_kとおくと,

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

となる。これはベルヌーイ分布の一般化である。

2. 多値変数

•あるデータ集合x_1 ,…,x_Nがとれたとき，尤度関数は，

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

となる。この式からN個のデータ点は，K個の

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

の値によってきまる。これは，各x_kが出た回数であり，十分統計量と呼ばれる。

この尤度関数を条件

$$\sum_{k=1}^{K}\mu_k = 1$$

のもとで最大化（ラグランジュ乗数法）し，最尤推定解を求めると，

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

となる。これは全体の中でx_kが出た割合。

2. 多項分布

•パラメータμ（各x_kが出る確率のベクトル）と観測値数Nが与えられたとき，$m_1 ,\ldots,m_K$の同時確率分布は

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

という多項分布の形になる。

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

は，N個のものを$m_1 ,\ldots,m_K$個ずつのK個のグループに分割する場合の数である。
したがって，次の制約条件が付く。

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

2．2．1 ディリクレ分布

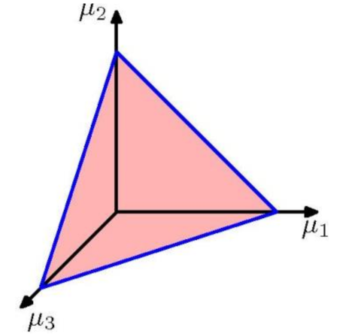
BehaviorInNetworks
都市生活行動学研究室

•多項分布もベイズ主義的に考える。
事前分布⇒多項分布と共役なもの

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

ただし，$0 \leqq \mu_k \leqq 1$, $\sum_k \mu_k = 1$

例）K=3のとき



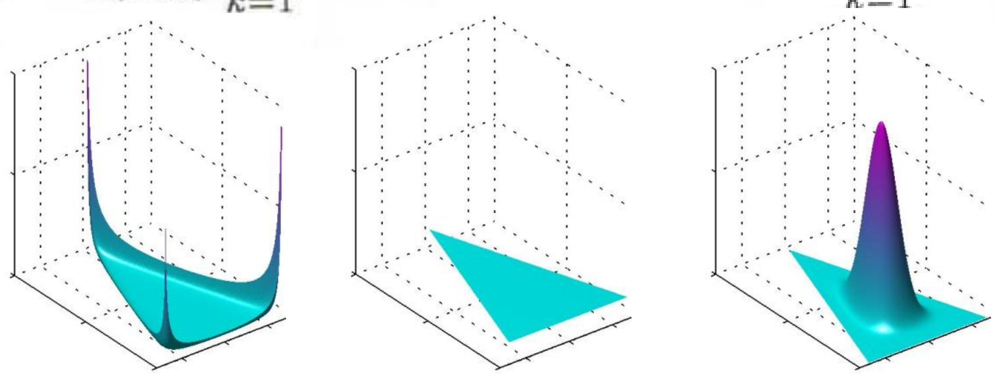
条件より，この分布はK-1次元
の単体上に制限される。

•この共役分布を正規化すると，次のディリクレ分布を得る

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

ただし $\alpha_0 = \sum_{k=1}^{K} \alpha_k$

例）K=3のとき
（縦軸を密度，横軸は単体上の座標）
左から
$\{\alpha_k\}=0.1$, $\{\alpha_k\}=1$, $\{\alpha_k\}=10$



2．2．1 ディリクレ分布

BehaviorInNetworks
都市生活行動学研究室

- 事後分布∝事前分布×尤度関数とすると。

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

という，事後分布もディリクレ分布の形になる。
結局，

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

となり，α_kはx_k =1となる有効観測数と解釈できる。

2.3 ガウス分布

BehaviorInNetworks
都市生活行動学研究室

- 1変数の場合

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

μは平均,σは分散

- 多変数の場合

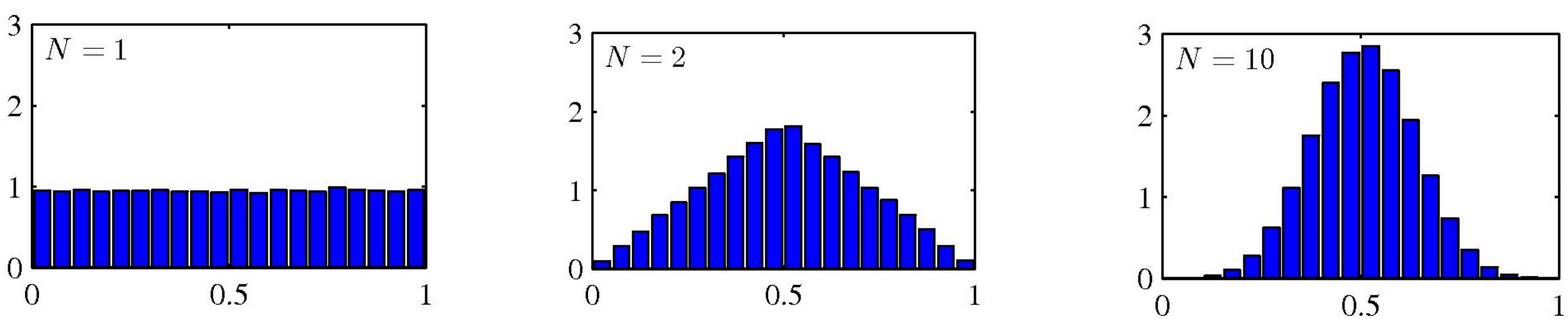
$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \quad (2.43)$$

μはD次元の平均ベクトル,∑はD×D共分散行列,|∑|は∑の行列式

- 中心極限定理

どんな分布に従う確率変数の和も，多くの和を取ると，ガウス分布に従うようになる。

例）一様分布に従うN個の確率変数の平均



2.3 ガウス分布の線形変換

BehaviorInNetworks
都市生活行動学研究室

•ガウス分布は

$$\Delta^2 = (\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \quad (2.44)$$

という二次形式でxに依存する。このΔをマハラノビス距離という。

ここで，共分散行列∑は実対称行列なので，i=1,…,Dについて固有方程式

$$\boldsymbol{\Sigma}\mathbf{u}_i = \lambda_i \mathbf{u}_i.$$

がかける。実対称行列は対角化できて，

対角成分がλ_iのベクトル →

$$\boldsymbol{\Sigma} = \sum_{i=1}^{D} \lambda_i \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}} \qquad \boldsymbol{\Sigma}^{-1} = \sum_{i=1}^{D} \frac{1}{\lambda_i} \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}}$$

となる。（2．44）に戻すと

$$\Delta^2 = \sum_{i=1}^{D} \frac{y_i^2}{\lambda_i} \quad (2.50)$$

ここで $y_i = \mathbf{u}_i^{\mathrm{T}}(\mathbf{x}-\boldsymbol{\mu}).$

となり，$\mathbf{y} = (y_1, \ldots, y_D)^{\mathrm{T}}$ にまとめると，

$$\mathbf{y} = \mathbf{U}(\mathbf{x}-\boldsymbol{\mu}) \quad (2.52)$$

となる。Uはu_i^{T}をi行に持つ直交行列。

2.3 ガウス分布の線形変換

•共分散行列の行列式|∑|も固有値の積でかけるので

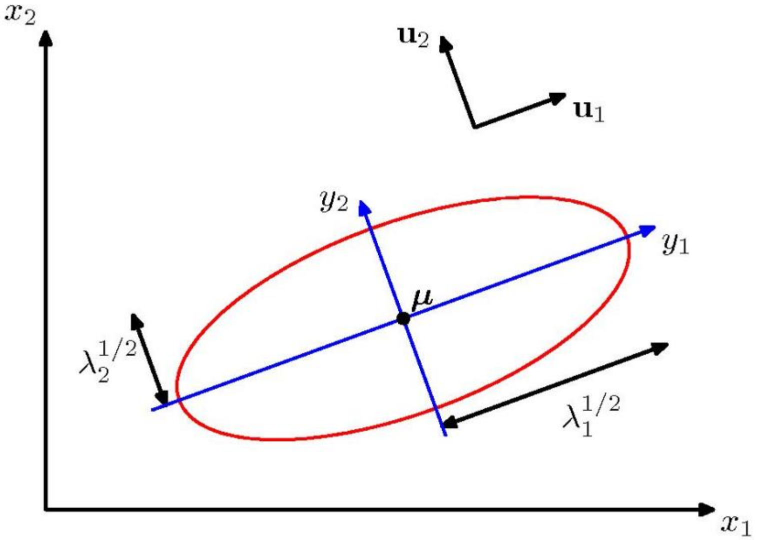
$$|\Sigma|^{1/2} = \prod_{j=1}^{D} \lambda_j^{1/2}$$

となり，（2.50）と合わせて(2.43)は

$$p(\mathbf{y}) = p(\mathbf{x})|\mathbf{J}| = \prod_{j=1}^{D} \frac{1}{(2\pi\lambda_j)^{1/2}} \exp\left\{-\frac{y_j^2}{2\lambda_j}\right\} \qquad (2.56)$$

と正規化できる。

例）二次元空間上のガウス分布
赤線内が密度一定の楕円体の面
(μ_1, μ_2)方向に新たに楕円の軸が定義される

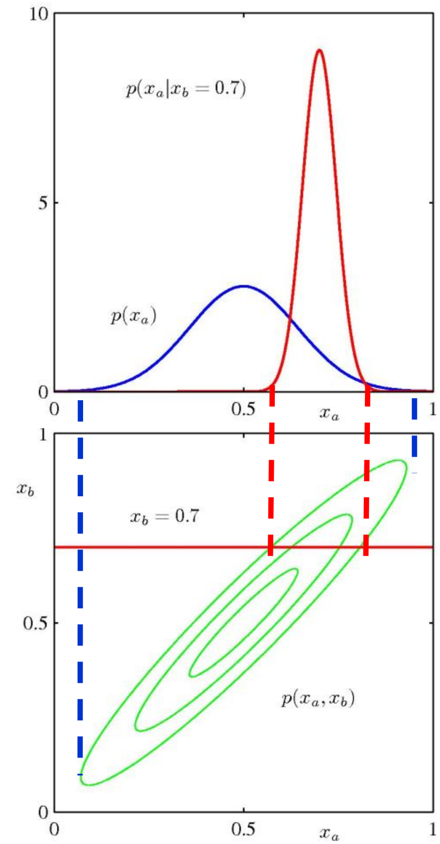


2.3.1 条件付きガウス分布

BehaviorInNetworks
都市生活行動学研究室

- 2つの変数集合の同時分布がガウス分布に従うとき
  - 一方の変数集合が与えられたとき，もう一方の条件付き分布もガウス分布に従う（赤）
  - どちらの変数集合の周辺分布もガウス分布に従う（青）C

例）二次元空間上のガウス分布
条件付き分布: x_b=0.7の軸で切った切り口
周辺分布：x_a軸から見た分布の正射影



2.3.1 条件付きガウス分布

BehaviorInNetworks
都市生活行動学研究室

•Xを二つの互いに素な部分集合X_aとX_bに分割する。平均と共分散行列も以下のように与えられる。

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{pmatrix} \quad (2.65) \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix} \quad (2.66) \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{pmatrix}. \quad (2.67)$$

ただし, $\boldsymbol{\Sigma}_{ba} = \boldsymbol{\Sigma}_{ab}^{\mathrm{T}}$ である。

•共分散行列の逆行列（対称行列になる）を考える。これを精度行列という。

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{aa} & \boldsymbol{\Lambda}_{ab} \\ \boldsymbol{\Lambda}_{ba} & \boldsymbol{\Lambda}_{bb} \end{pmatrix}.$$

同様に, $\boldsymbol{\Lambda}_{ba}^{\mathrm{T}} = \boldsymbol{\Lambda}_{ab}$ である。

2.3.1 条件付きガウス分布

•$p(x_a|x_b)$の表現を考えるため，ガウス分布の指数部分に着目する。指数部分の二次形式は(2.65),(2.66)より，

$$
\begin{aligned}
-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = \\
-\frac{1}{2}(\mathbf{x}_a-\boldsymbol{\mu}_a)^{\mathrm{T}}\boldsymbol{\Lambda}_{aa}(\mathbf{x}_a-\boldsymbol{\mu}_a) - \frac{1}{2}(\mathbf{x}_a-\boldsymbol{\mu}_a)^{\mathrm{T}}\boldsymbol{\Lambda}_{ab}(\mathbf{x}_b-\boldsymbol{\mu}_b) \\
-\frac{1}{2}(\mathbf{x}_b-\boldsymbol{\mu}_b)^{\mathrm{T}}\boldsymbol{\Lambda}_{ba}(\mathbf{x}_a-\boldsymbol{\mu}_a) - \frac{1}{2}(\mathbf{x}_b-\boldsymbol{\mu}_b)^{\mathrm{T}}\boldsymbol{\Lambda}_{bb}(\mathbf{x}_b-\boldsymbol{\mu}_b) \qquad (2.70)
\end{aligned}
$$

と分解でき，X_bを固定して考えると，X_a の二次形式になっている。これと，次の一般のXの二次形式と比較する。

$$
-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = -\frac{1}{2}\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x} + \mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \text{const.} \qquad (2.71)
$$

2.3.1 条件付きガウス分布

BehaviorInNetworks
都市生活行動学研究室

- (2.70) のx_aの二次の項をとりだすと,

$$-\frac{1}{2}\mathbf{x}_a^{\mathrm{T}}\mathbf{\Lambda}_{aa}\mathbf{x}_a$$

これを(2.71)のx_aの二次の項と比較して

$$\mathbf{\Sigma}_{a|b} = \mathbf{\Lambda}_{aa}^{-1}$$

を得る。次に一次の項を考え, $\mathbf{\Lambda}_{ba}^{\mathrm{T}} = \mathbf{\Lambda}_{ab}$ の性質を利用すると,

$$\mathbf{x}_a^{\mathrm{T}}\left\{\mathbf{\Lambda}_{aa}\boldsymbol{\mu}_a - \mathbf{\Lambda}_{ab}(\mathbf{x}_b - \boldsymbol{\mu}_b)\right\}$$

という一次の項が得られる。これと一般形(2.71)の一次の項を比較して,

$$\begin{aligned}\boldsymbol{\mu}_{a|b} &= \mathbf{\Sigma}_{a|b}\left\{\mathbf{\Lambda}_{aa}\boldsymbol{\mu}_a - \mathbf{\Lambda}_{ab}(\mathbf{x}_b - \boldsymbol{\mu}_b)\right\} \\ &= \boldsymbol{\mu}_a - \mathbf{\Lambda}_{aa}^{-1}\mathbf{\Lambda}_{ab}(\mathbf{x}_b - \boldsymbol{\mu}_b)\end{aligned}$$

という形で平均が得られる。

2.3.1 条件付きガウス分布

$$\begin{aligned}\boldsymbol{\mu}_{a|b} &= \boldsymbol{\Sigma}_{a|b}\{\boldsymbol{\Lambda}_{aa}\boldsymbol{\mu}_a - \boldsymbol{\Lambda}_{ab}(\mathbf{x}_b - \boldsymbol{\mu}_b)\} \\ &= \boldsymbol{\mu}_a - \boldsymbol{\Lambda}_{aa}^{-1}\boldsymbol{\Lambda}_{ab}(\mathbf{x}_b - \boldsymbol{\mu}_b)\end{aligned}$$

$$\boldsymbol{\Sigma}_{a|b} = \boldsymbol{\Lambda}_{aa}^{-1}$$

をもとの分割された共分散行列で表現する。関形式

$$\begin{pmatrix}\boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb}\end{pmatrix}^{-1} = \begin{pmatrix}\boldsymbol{\Lambda}_{aa} & \boldsymbol{\Lambda}_{ab} \\ \boldsymbol{\Lambda}_{ba} & \boldsymbol{\Lambda}_{bb}\end{pmatrix}$$

と，次のシューア補行列を用いて，

$$\begin{pmatrix}\mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D}\end{pmatrix}^{-1} = \begin{pmatrix}\mathbf{M} & -\mathbf{M}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{M} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}\mathbf{M}\mathbf{B}\mathbf{D}^{-1}\end{pmatrix}.$$

$$\mathbf{M} = (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}.$$

$$\begin{aligned}\boldsymbol{\Lambda}_{aa} &= (\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba})^{-1} \\ \boldsymbol{\Lambda}_{ab} &= -(\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba})^{-1}\boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\end{aligned}$$

という関形式を得るので，結局，

$$\boxed{\begin{aligned}\boldsymbol{\mu}_{a|b} &= \boldsymbol{\mu}_a + \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}(\mathbf{x}_b - \boldsymbol{\mu}_b) \\ \boldsymbol{\Sigma}_{a|b} &= \boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba}\end{aligned}}$$

となり，条件付きガウス分布の平均と分散が表現できる。

2.3.2 周辺ガウス分布

•周辺分布

$$p(\mathbf{x}_a) = \int p(\mathbf{x}_a, \mathbf{x}_b)\,\mathrm{d}\mathbf{x}_b$$

もガウス分布になることを確認する。

まず，X_bを積分消去するため，(2.70)からX_bを含む項を取り出すと，

$$-\frac{1}{2}\mathbf{x}_b^{\mathrm{T}}\boldsymbol{\Lambda}_{bb}\mathbf{x}_b + \mathbf{x}_b^T\mathbf{m} = -\frac{1}{2}(\mathbf{x}_b - \boldsymbol{\Lambda}_{bb}^{-1}\mathbf{m})^{\mathrm{T}}\boldsymbol{\Lambda}_{bb}(\mathbf{x}_b - \boldsymbol{\Lambda}_{bb}^{-1}\mathbf{m}) + \frac{1}{2}\mathbf{m}^{\mathrm{T}}\boldsymbol{\Lambda}_{bb}^{-1}\mathbf{m} \quad (2.84)$$

ただし， $\mathbf{m} = \boldsymbol{\Lambda}_{bb}\boldsymbol{\mu}_b - \boldsymbol{\Lambda}_{ba}(\mathbf{x}_a - \boldsymbol{\mu}_a)$

（２．８４）の右辺第１項は標準的なガウス分布の二次形式部分なので，この部分を指数にとった積分

$$\int \exp\left\{-\frac{1}{2}(\mathbf{x}_b - \boldsymbol{\Lambda}_{bb}^{-1}\mathbf{m})^{\mathrm{T}}\boldsymbol{\Lambda}_{bb}(\mathbf{x}_b - \boldsymbol{\Lambda}_{bb}^{-1}\mathbf{m})\right\}\mathrm{d}\mathbf{x}_b.$$

は共分散行列の逆行列にのみ依存する正規化係数の逆数になる。
よってX_bを積分消去することができる。

2.3.2 周辺ガウス分布

BehaviorInNetworks
都市生活行動学研究室

•mも含め，残るX_aに関する項のみをまとめると，

$$\frac{1}{2}\left[\Lambda_{bb}\boldsymbol{\mu}_b - \Lambda_{ba}(\mathbf{x}_a - \boldsymbol{\mu}_a)\right]^{\mathrm{T}} \Lambda_{bb}^{-1}\left[\Lambda_{bb}\boldsymbol{\mu}_b - \Lambda_{ba}(\mathbf{x}_a - \boldsymbol{\mu}_a)\right]$$
$$-\frac{1}{2}\mathbf{x}_a^{\mathrm{T}}\Lambda_{aa}\mathbf{x}_a + \mathbf{x}_a^{\mathrm{T}}(\Lambda_{aa}\boldsymbol{\mu}_a + \Lambda_{ab}\boldsymbol{\mu}_b) + \mathrm{const}$$
$$= -\frac{1}{2}\mathbf{x}_a^{\mathrm{T}}(\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba})\mathbf{x}_a$$
$$+\mathbf{x}_a^{\mathrm{T}}(\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba})\boldsymbol{\mu}_a + \mathrm{const}$$

これを再び(2.71)と比較すると，

$$\Sigma_a = (\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba})^{-1}$$

これより，平均は

$$\Sigma_a(\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba})\boldsymbol{\mu}_a = \boxed{\boldsymbol{\mu}_a}$$

となり，精度行列とシューア補行列を用いて，

$$(\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba})^{-1} = \boxed{\Sigma_{aa}}$$

となるので，結局，周辺分布の平均と分散は，それぞれの集合の平均と分散であらわされることがわかる。

2.3.3 ガウス変数に対するベイズの定理

BehaviorInNetworks
都市生活行動学研究室

•あるガウス周辺分布p(x)と平均がxの線形関数で，共分散はxとは独立なガウス条件付き分布p(y|x)を考える。
これは，線形ガウスモデルの例である。

周辺分布と条件付き分布を

$$p(\mathbf{x}) = \mathcal{N}\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1}\right)$$

$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}\left(\mathbf{y}|\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{L}^{-1}\right)$$

とおく。

2.3.3 ガウス変数に対するベイズの定理

•まず，xとyの同時分布の表現を求める。次のようなzを定義する。

$$\mathbf{z} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$$

同時分布の対数を考えると，

$$\begin{aligned} \ln p(\mathbf{z}) &= \ln p(\mathbf{x}) + \ln p(\mathbf{y}|\mathbf{x}) \\ &= -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Lambda}(\mathbf{x}-\boldsymbol{\mu}) \\ &\quad -\frac{1}{2}(\mathbf{y}-\mathbf{A}\mathbf{x}-\mathbf{b})^{\mathrm{T}}\mathbf{L}(\mathbf{y}-\mathbf{A}\mathbf{x}-\mathbf{b}) + \text{const} \end{aligned}$$

x,yの2次の項を取り出すと，

$$\begin{aligned} &-\frac{1}{2}\mathbf{x}^{\mathrm{T}}(\boldsymbol{\Lambda}+\mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{A})\mathbf{x} - \frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{L}\mathbf{y} + \frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{L}\mathbf{A}\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{y} \\ &= -\frac{1}{2}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}^{\mathrm{T}} \boxed{\begin{pmatrix} \boldsymbol{\Lambda}+\mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{A} & -\mathbf{A}^{\mathrm{T}}\mathbf{L} \\ -\mathbf{L}\mathbf{A} & \mathbf{L} \end{pmatrix}} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = -\frac{1}{2}\mathbf{z}^{\mathrm{T}}\boxed{\mathbf{R}}\mathbf{z}. \end{aligned}$$

よって，共分散行列は

$$\boxed{\operatorname{cov}[\mathbf{z}] = \mathbf{R}^{-1} = \begin{pmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{\Lambda}^{-1}\mathbf{A}^{\mathrm{T}} \\ \mathbf{A}\boldsymbol{\Lambda}^{-1} & \mathbf{L}^{-1}+\mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^{\mathrm{T}} \end{pmatrix}}$$

2.3.3 ガウス変数に対するベイズの定理

- 平均は1次の項

$$\mathbf{x}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{\mu} - \mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{b} + \mathbf{y}^{\mathrm{T}}\mathbf{L}\mathbf{b} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \boldsymbol{\Lambda}\boldsymbol{\mu} - \mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{b} \\ \mathbf{L}\mathbf{b} \end{pmatrix}$$

と $\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ を比較して，

$$\boxed{\mathbb{E}[\mathbf{z}] = \mathbf{R}^{-1} \begin{pmatrix} \boldsymbol{\Lambda}\boldsymbol{\mu} - \mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{b} \\ \mathbf{L}\mathbf{b} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{A}\boldsymbol{\mu} + \mathbf{b} \end{pmatrix}}$$

を得る。また，これと，

$$\begin{pmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{\Lambda}^{-1}\mathbf{A}^{\mathrm{T}} \\ \mathbf{A}\boldsymbol{\Lambda}^{-1} & \mathbf{L}^{-1} + \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^{\mathrm{T}} \end{pmatrix} \quad と \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{pmatrix}.$$

の比較により，yの周辺確率の平均と分散

$$\mathbb{E}[\mathbf{y}] = \mathbf{A}\boldsymbol{\mu} + \mathbf{b}$$

$$\mathrm{cov}[\mathbf{y}] = \mathbf{L}^{-1} + \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^{\mathrm{T}}$$

が得られる。

2.4.4 ガウス分布の最尤推定

BehaviorInNetworks
都市生活行動学研究室

•ある多変量ガウス分布から，観測値$\{x_n\}$が独立に取れたデータ集合$X=(x_1,\ldots,x_N)^T$がある時，パラメータを最尤推定法で求める。
•対数尤度関数は，

$$\ln p(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{n=1}^{N}(\mathbf{x}_n-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n-\boldsymbol{\mu}) \quad (2.118)$$

なので，これを最大化する。μ,Σで偏微分し，=0とおいて， μ,Σの最尤推定量を求めると，

$$\boxed{\boldsymbol{\mu}_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n.} \qquad \boldsymbol{\Sigma}_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n-\boldsymbol{\mu}_{\mathrm{ML}})(\mathbf{x}_n-\boldsymbol{\mu}_{\mathrm{ML}})^{\mathrm{T}}.$$

この真の分布の下での期待値を求めると，

$$\mathbb{E}[\boldsymbol{\mu}_{\mathrm{ML}}] = \boldsymbol{\mu}$$

$$\mathbb{E}[\boldsymbol{\Sigma}_{\mathrm{ML}}] = \frac{N-1}{N}\boldsymbol{\Sigma}$$

となるが，共分散は不偏推定量ではないので，次のように補正する。

$$\boxed{\widetilde{\boldsymbol{\Sigma}} = \frac{1}{N-1}\sum_{n=1}^{N}(\mathbf{x}_n-\boldsymbol{\mu}_{\mathrm{ML}})(\mathbf{x}_n-\boldsymbol{\mu}_{\mathrm{ML}})^{\mathrm{T}}.}$$

逐次推定は来週