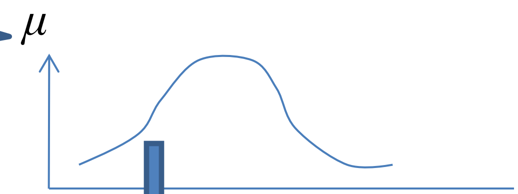
2. Bayes統計に基づく推論

Bayesによる確率分布推定の考え方
多項分布、ディリクレ分布
事前分布としてのディリクレ分布の意味
正規分布と事後分布
多次元正規分布と条件付き分布
指数型分布族
自然共役事前分布の最尤推定

by 中川裕志（東京大学）

Bayesによる確率分布推定の考え方

事前分布 とはパラメター
（i.e. μ）自体の分布

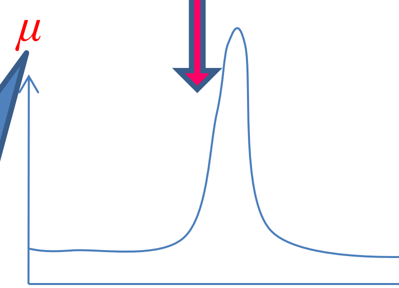


観測データ
or 教師データ:X

$p(\mu|X) = p(X|\mu)\ p(\mu)$

観測データを事前分布にBayesの定理で組み合わせる

Xを観測した後に得たパラメターμの事後分布



パラメター μは点ではなく、分布として与えられる点に注意！

多項分布:$Mult$

- 複数の離散データが独立に出現する場合の確率分布の定番
- 個々の離散データ間に相関がない場合に使うもので基本的分布。
  - 以下はK種類の離散データ(例えば、語彙数がKでN単語からなるテキストでの単語の出現分布)がある場合

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

$$\mathrm{E}\left[m_k\right] = N\mu_k \qquad \mathrm{var}\left[m_k\right] = N\mu_k\left(1-\mu_k\right) \qquad \mathrm{cov}\left[m_j m_k\right] = -N\mu_j\mu_k$$

ただし、 $\sum_{i=1}^{K} \mu_k = 1$

ディリクレ分布: Dir

- 多項分布では離散事象（たとえば単語）iの出現回数 m_i が確率変数だった。
- しかし、逆に m_i が観測値として既知の場合に、単語 i の出現確率 μ_i が確率変数となる分布も考えられる。すなわち、多項分布の事前分布として使えるような分布。

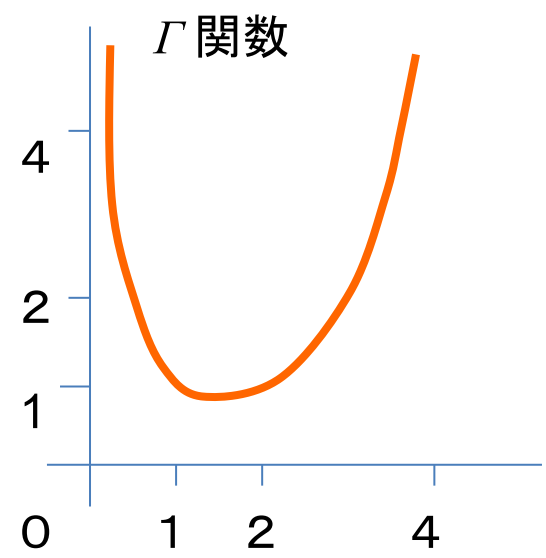
ディリクレ分布：Dir

- K変数の場合。αはパラメターだが、以下の式の分布を作るときに使った既知の観測回数のデータと考えてもよいだろう。

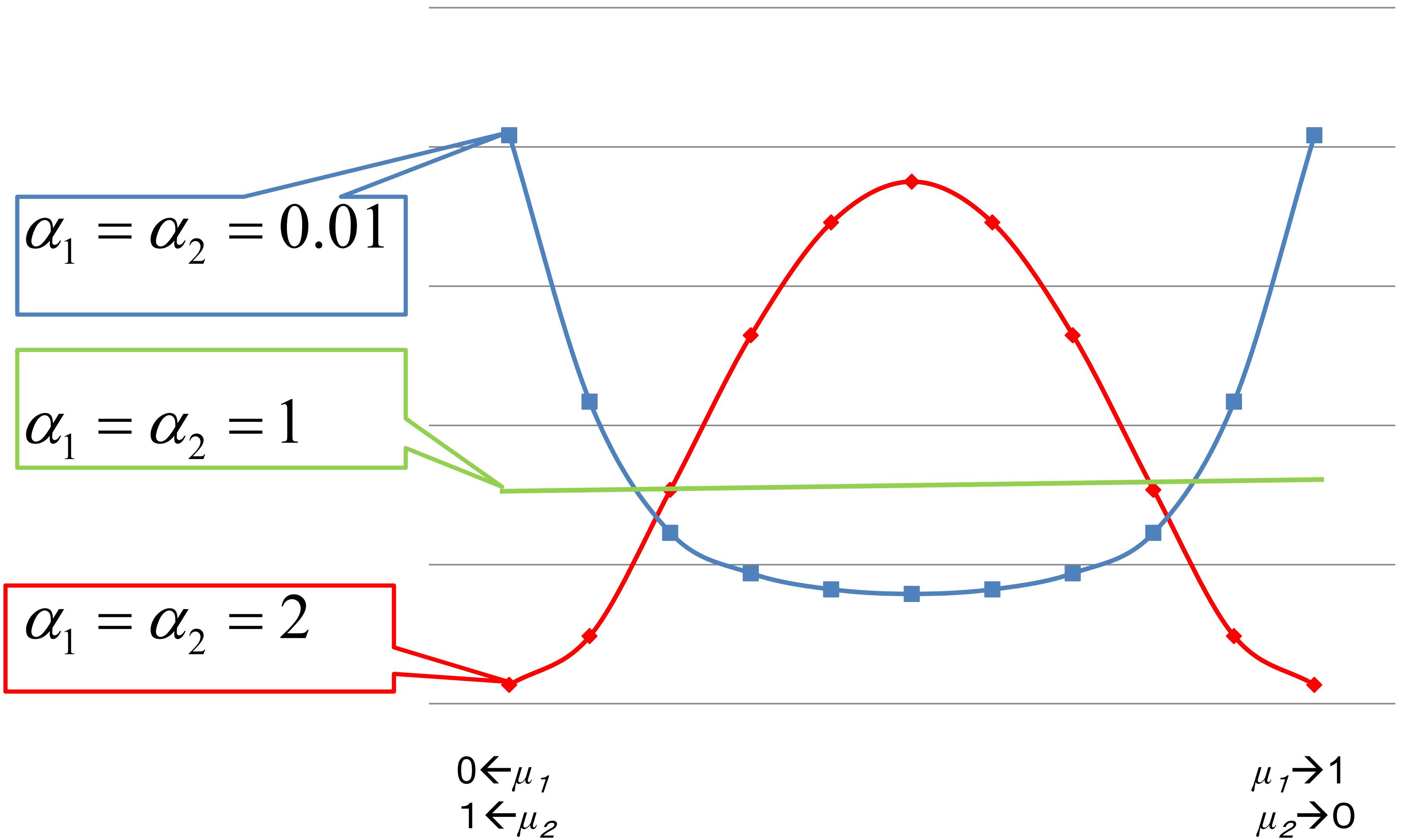
$$\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_K \end{pmatrix} \qquad \alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_K \end{pmatrix} \qquad \sum_{k=1}^{K} \mu_k = 1 \qquad 0 \le \mu_k \le 1$$

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

$$\mathrm{E}[\mu_k] = \frac{\alpha_k}{\alpha_1 + \cdots + \alpha_K}$$



ディリクレ分布の例



$\alpha_1 = \alpha_2 = 0.01$
$\alpha_1 = \alpha_2 = 1$
$\alpha_1 = \alpha_2 = 2$
0←μ_1
1←μ_2
μ_1→1
μ_2→0

事前分布としてのディリクレ分布の意味

- ディリクレ分布Dirを事前分布とみなして、観測データが多項分布Multで与えられたときの事後分布としてのディリクレ分布Dirを考える

観測データ　iの出現回数　m_i: $X=(m_1,\ldots,m_K)$,　$\sum_{i=1}^{K} m_i = M, \sum_{i=1}^{K} \alpha_i = \alpha_0$

事後　観測　事前

$$Dir(\mu \mid X,\alpha) \propto Mult(X \mid \mu) Dir(\mu \mid \alpha) \propto \prod_{i=1} \mu_i^{\alpha_i + m_i - 1}$$

$$\Rightarrow \quad \boxed{Dir(\mu \mid X,\alpha) = Dir(\mu \mid \alpha + X) = \frac{\Gamma(\alpha_0 + M)}{\Gamma(\alpha_1 + m_1)\cdots\Gamma(\alpha_K + m_K)} \prod_{i=1}^{K} \mu_i^{\alpha_i + m_i - 1}}$$

- こうして見ると、α_iは事前分布を得るために想定したiの（仮想的）観測回数と見做せる。

正規分布(1変数)と事後分布

- 1変数正規分布:連続する数値データの確率分布の定番

$$N(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \qquad \mathrm{E}[x] = \mu, \qquad \mathrm{var}[x] = \sigma^2$$

- では、ここでMultからDirの事後分布を求めたと同じように、Bayesの定理を用いて、正規分布において、事前分布から事後分布を求めてみよう。
  - 次のページの例は簡単のため、分散は既知とし、事後分布の期待値だけを求めることにする。
    - 分散の事後分布についてはWishart分布という分布が登場するが、難しいのでここでは省略

事前分布 : $p(\mu) = N(\mu \mid \mu_0, \sigma_0^2)$

K個の観測データが得られた場合の尤度 : $p(X \mid \mu)$

$$= \prod_{i=1}^{K} p(x_i \mid \mu) = \frac{1}{(2\pi\sigma^2)^{K/2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{K} (x_i - \mu)^2 \right\}$$

ただしσは既知

Bayesの定理から

事後 観測 事前

$$p(\mu \mid X) \propto p(X \mid \mu)\, p(\mu) \propto \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{K} (x_i - \mu)^2 - \frac{(\mu - \mu_0)^2}{2\sigma_0^2} \right\}$$

$$\propto \exp\left\{ -\frac{\frac{K}{\sigma^2} + \frac{1}{\sigma_0^2}}{2} \mu^2 + \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^{K} x_i}{\sigma^2} \right) \mu + \cdots \right\} \quad -\boxed{(N10)}$$

$p(\mu \mid X)$は正規分布だからこの結果より

$\Rightarrow$ K個の観測データを得た後の事後分布 : $p(\mu \mid X) = N(\mu \mid \mu_K, \sigma_K^2)$

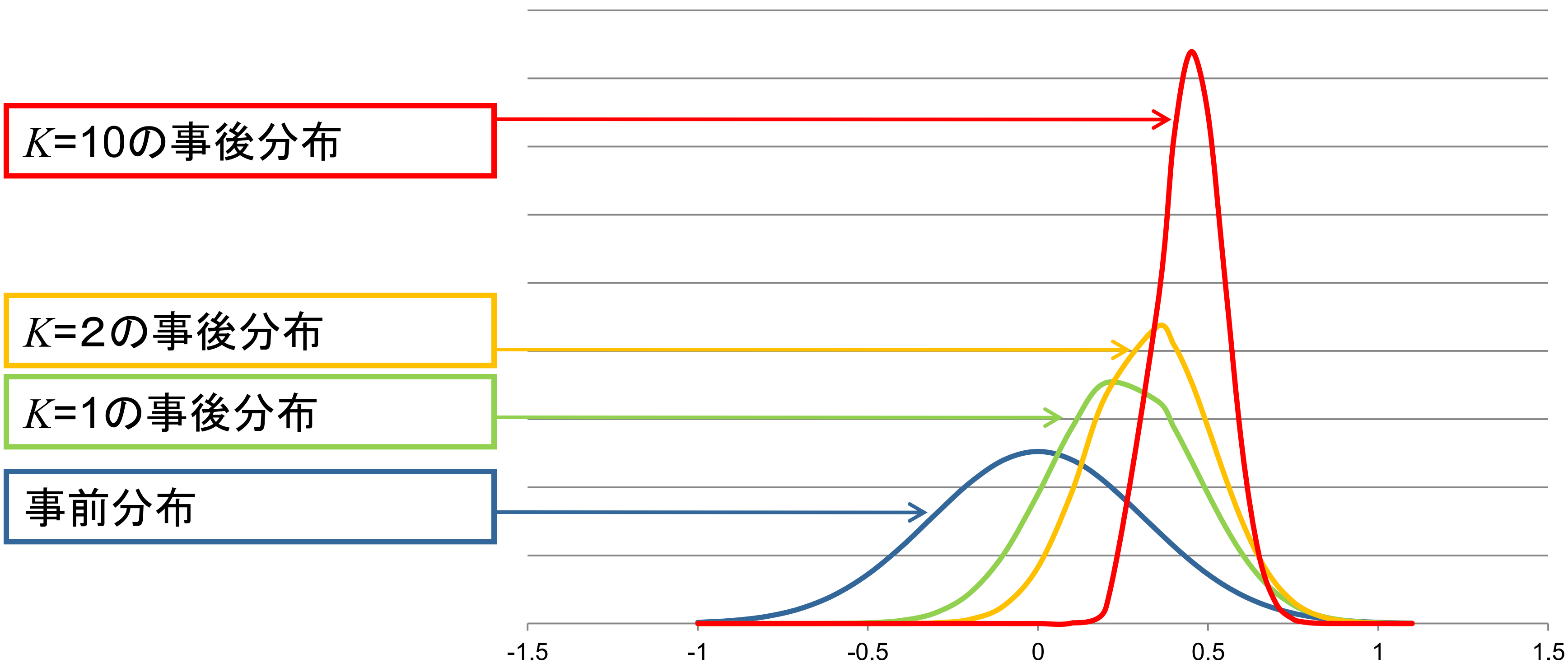
$$\frac{1}{\sigma_K^2} = \frac{1}{\sigma_0^2} + \frac{K}{\sigma^2}, \qquad \mu_K = \frac{\sigma^2}{K\sigma_0^2 + \sigma^2} \mu_0 + \frac{\sigma_0^2}{K\sigma_0^2 + \sigma^2} \sum_{i=1}^{K} x_i$$

事前分布からの寄与

観測データからの寄与

観測データ数Kと事後分布の例

$$\sigma_0^2 = 0.1 \qquad \sigma^2 = 0.1 \qquad \mu_0 = 0 \qquad \mu = \frac{1}{K}\sum_{i=1}^{K} x_i \qquad E[x] = 0.5$$



観測データにより事前分布の
パラメータμが修正されていく

多次元正規分布

- 多次元正規分布：複数種類（つまり複数の確率変数）を持つ数値データの確率分布

D次元の正規分布

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_D \end{bmatrix} \qquad \mathrm{E}[\mathbf{x}] = \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_D \end{bmatrix}$$

$$\text{共分散行列 :}\mathrm{cov}[\mathbf{x}] = \boldsymbol{\Sigma} = \begin{bmatrix} {\sigma^2}_1 & & \sigma_{1D} \\ & \ddots & \\ \sigma_{D1} & & {\sigma^2}_D \end{bmatrix} = \Lambda^{-1} : \text{精度行列}$$

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

$$= \frac{|\Lambda|^{1/2}}{(2\pi)^{D/2}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Lambda (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

例題

- 多次元正規分布の共分散行列を推定する。

D次元の正規分布

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

$$\log N(\mathbf{x}\,|\,\boldsymbol{\mu},\Lambda^{-1})\propto\log|\Lambda|-(\mathbf{x}-\boldsymbol{\mu})^T\Lambda(\mathbf{x}-\boldsymbol{\mu})$$

この対数尤度を最大化するためにΛで微分して0とおく

第1項の微分 $$\frac{\partial\log|\Lambda|}{\partial\Lambda}=\left(\Lambda^{-1}\right)^T=\Lambda^{-1}$$

第2項の微分 $$\frac{\partial(\mathbf{x}-\boldsymbol{\mu})^T\Lambda(\mathbf{x}-\boldsymbol{\mu})}{\partial\Lambda}=\frac{\partial trace\left(\Lambda(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\right)}{\partial\Lambda}$$

$$=\left((\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\right)^T=(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T$$

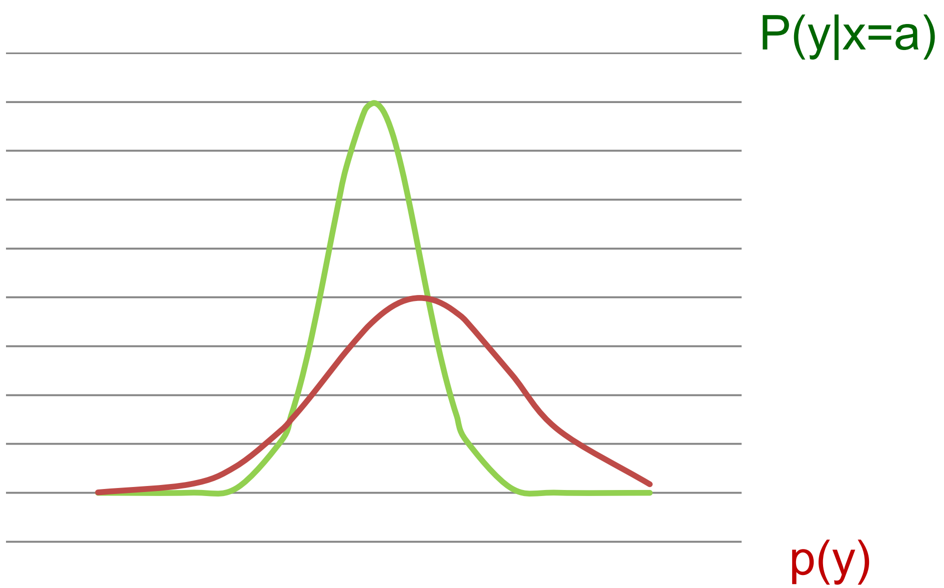
ここでxに対して期待値をとる、すなわち$E_{\mathrm{x}}[]$をすると

$$\Lambda^{-1}-E_{\mathrm{x}}[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T]=0$$

より$E_{\mathrm{x}}[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T]=\Lambda^{-1}=\Sigma$

条件付正規分布

- 変数ベクトルzをxとyに分割すると



- 変数ベクトルzをxとyに分割する。

多次元正規分布 $N(\mathbf{z}|\boldsymbol{\mu},\Sigma)$

$$\mathbf{z}=\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}\quad \boldsymbol{\mu}=\begin{pmatrix}\boldsymbol{\mu}_x\\ \boldsymbol{\mu}_y\end{pmatrix}\quad \Sigma=\begin{pmatrix}\Sigma_{xx} & \Sigma_{xy}\\ \Sigma_{yx} & \Sigma_{yy}\end{pmatrix}$$

$$where\quad \Sigma=\Sigma^T \qquad \Sigma_{xy}=\Sigma_{yx}{}^T$$

精度行列：$\Lambda\equiv\Sigma^{-1}$　とすると　$\Lambda=\begin{pmatrix}\Lambda_{xx} & \Lambda_{xy}\\ \Lambda_{yx} & \Lambda_{yy}\end{pmatrix}$　$\Lambda_{xy}{}^T=\Lambda_{yx}$

- ここで多次元正規分布の指数の肩の項は次式

$$\boxed{\begin{aligned}&-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu})=\\ &\quad -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_x)^T\Lambda_{xx}(\mathbf{x}-\boldsymbol{\mu}_x)-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_x)^T\Lambda_{xy}(\mathbf{y}-\boldsymbol{\mu}_y)\\ &\quad -\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu}_y)^T\Lambda_{yx}(\mathbf{x}-\boldsymbol{\mu}_x)-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu}_y)^T\Lambda_{yy}(\mathbf{y}-\boldsymbol{\mu}_y)\end{aligned}}\quad \text{-(G-10)}$$

➢一般に正規分布 $N(\mathbf{z}\,|\,\boldsymbol{\mu},\Sigma)$ の指数の肩は次式で書け、右辺の第１項、第２項の係数の部分に着目すれば期待値、共分散が求まる。

$$-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu}) = -\frac{1}{2}\mathbf{z}^T\Sigma^{-1}\mathbf{z} + \mathbf{z}^T\Sigma^{-1}\boldsymbol{\mu} + const$$

-（G-20)

➢条件付正規分布***p(x|y)***の期待値$\mu_{x|y}$と共分散$\Sigma_{x|y}$をこの方法を（G−10)式に適用して求めよう。ー　問題

- yを定数とみなしてxの分布を求めれば、条件付分布になるから(G-10)の第1項のxの2次の項の係数が共分散。すなわち

$$\boxed{-\frac{1}{2}\mathbf{x}^T\Lambda_{xx}\mathbf{x} \quad \text{により} \quad \Sigma_{x|y}=\Lambda_{xx}{}^{-1}}$$

- 一方、(G-10)において$\mathbf{x}$の1次の項が$\Sigma^{-1}\boldsymbol{\mu}$　これは次式

$$\mathbf{x}^T\left\{\Lambda_{xx}\boldsymbol{\mu}_x-\Lambda_{xy}(\mathbf{y}-\boldsymbol{\mu}_y)\right\}$$

これにより

$$\boxed{\begin{aligned}\boldsymbol{\mu}_{x|y}&=\Sigma_{x|y}\left\{\Lambda_{xx}\boldsymbol{\mu}_x-\Lambda_{xy}(\mathbf{y}-\boldsymbol{\mu}_y)\right\} \quad \Sigma_{x|y}=\Lambda_{xx}{}^{-1}\text{より}\\ &=\boldsymbol{\mu}_x-\Lambda_{xx}{}^{-1}\Lambda_{xy}(\mathbf{y}-\boldsymbol{\mu}_y)\end{aligned}}$$

- 次に、これらの結果を共分散行列を用いて書き直す

$$\begin{pmatrix}\Sigma_{xx} & \Sigma_{xy}\\ \Sigma_{yx} & \Sigma_{yy}\end{pmatrix}^{-1}=\begin{pmatrix}\Lambda_{xx} & \Lambda_{xy}\\ \Lambda_{yx} & \Lambda_{yy}\end{pmatrix} \quad \text{において} \quad (Matrix^{-1})\text{を使えば}$$

$$\Lambda_{xx}=(\Sigma_{xx}-\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})^{-1} \qquad \Lambda_{xy}=-(\Sigma_{xx}-\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}$$

$$\Rightarrow \boxed{\begin{aligned}\boldsymbol{\mu}_{x|y}&=\boldsymbol{\mu}_x+\Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y}-\boldsymbol{\mu}_y)\\ \Sigma_{x|y}&=\Sigma_{xx}-\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\end{aligned}}$$

exponential family:指数型分布族

➢事前分布と学習後の事後分布が同一タイプの分布(事前共役)

$$p(\mathbf{x}\mid\eta)=h(\mathbf{x})\exp(\eta^T u(\mathbf{x})-a(\eta)) \qquad (EB1)$$

ただし,$x,u(\mathbf{x}),\eta$は一般にはベクトル

また、$\exp(-a(\eta))\int h(\mathbf{x})\exp(\eta^T u(\mathbf{x}))\mathrm{d}\mathbf{x}=1$

$$a(\eta)=\log\left(\int h(x)\exp(\eta^T u(x))\mathrm{d}x)\right) \qquad (EB2)$$

正規化項

iidの観測データ$X=\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$に対しては以下の式

$$p(\mathbf{X}\mid\eta)=\left(\prod_{n=1}^{N}h(\mathbf{x}_n)\right)\exp\left(\eta^T\sum_{n=1}^{N}\mathbf{u}(\mathbf{x}_n)-Na(\eta)\right)$$

いくつかの確率密度関数のExponential family表現：ガウス分布

$$p(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{x^2 - 2x\mu + \mu^2}{2\sigma^2} \right\}$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left\{ \underbrace{\left(\frac{\mu}{\sigma^2}, \frac{1}{\sigma^2} \right)}_{(\eta_1,\ \eta_2) = \eta^T} \underbrace{\begin{pmatrix} x \\ -\dfrac{x^2}{2} \end{pmatrix}}_{x} - \underbrace{\frac{\mu^2}{2\sigma^2} - \log\sigma}_{-\left(\frac{\eta_1^2}{2\eta_2} + \log\sigma \right) = -a(\eta)} \right\}$$

いくつかの確率密度関数のExponential family表現：多項分布

多項分布($Multinomial$)のexponential family表現

$$p(x|\mu)=\binom{N}{x_1x_2\cdots x_K}\prod_{k=1}^{K}\mu_k^{x_k}\qquad \sum_{k=1}^{K}\mu_k=1,\sum_{k=1}^{K}x_k=N \text{を使うと}$$

$$=\binom{N}{x_1x_2\cdots x_K}\exp\left\{\sum_{k=1}^{K-1}x_k\log\mu_k+\left(N-\sum_{k=1}^{K-1}x_k\right)\left(\log\left(1-\sum_{k=1}^{K-1}\mu_k\right)\right)\right\}$$

$$=\underbrace{\binom{N}{x_1x_2\cdots x_K}}_{h(x)}\exp\left\{\underbrace{\sum_{k=1}^{K-1}\log\left(\frac{\mu_k}{1-\sum_{k=1}^{K-1}\mu_k}\right)}_{\mu^T}\underbrace{x_k}_{x}+\underbrace{N\left(\log\left(1-\sum_{k=1}^{K-1}\mu_k\right)\right)}_{a(\mu)}\right\}$$

自然共役事前分布の最尤推定

$$p(\mathbf{x}\,|\,\eta)=h(\mathbf{x})\exp(\eta^T u(\mathbf{x})-a(\eta))$$

におけるnatural parameter ηの最尤推定をする。

$$1=\int p(\mathbf{x}\,|\,\eta)d\mathbf{x}=\int h(\mathbf{x})\exp(\eta^T u(\mathbf{x})-a(\eta))d\mathbf{x}$$

上の式をηの第j成分η_jで微分してゼロとおくと

$$-\frac{\partial a(\eta)}{\partial \eta_j}\underbrace{\int h(\mathbf{x})\exp(\eta^T u(\mathbf{x})-a(\eta))d\mathbf{x}}_{=1}+\underbrace{\int h(\mathbf{x})\exp(\eta^T u(\mathbf{x})-a(\eta))u(\mathbf{x})_j}_{=E[\mathbf{u}(\mathbf{x})_j]}d\mathbf{x}=0$$

$$\frac{\partial a(\eta)}{\partial \eta_j}=E[\mathbf{u}(\mathbf{x})_j] \qquad (EB3)$$

$$\frac{\partial^2 a(\eta)}{\partial {\eta_j}^2}=V[\mathbf{u}(\mathbf{x})_j] \qquad (EB4)$$

(EB3)(EB4)の応用例

➢ガウス分布に応用

$$p(x \mid \mu, \sigma^2) == \frac{1}{\sqrt{2\pi}} \exp\left\{ \left(\frac{\mu}{\sigma^2}, \frac{1}{\sigma^2} \right) \begin{pmatrix} x \\ -\frac{x^2}{2} \end{pmatrix} - \frac{\mu^2}{2\sigma^2} - \log\sigma \right\}$$

$$\eta^T = (\eta_1,\ \eta_2) \quad x \qquad a(\eta) = -\left(\frac{\eta_1{}^2}{2\eta_2} - \frac{1}{2}\log\eta_2 \right)$$

$$E_{\eta 1}[x] = \frac{\partial a(\eta)}{\partial \eta_1} = \frac{\partial}{\partial \eta_1} \frac{\eta_1{}^2}{2\eta_2} = \frac{\eta_1}{\eta_2} = \mu \qquad V_{\eta 1}[x] = \frac{\partial^2 a(\eta)}{\partial \eta_1{}^2} = \frac{\partial}{\partial \eta_1} \frac{\eta_1}{\eta_2} = \frac{1}{\eta_2} = \sigma^2$$

(EB3)(EB4)の応用例　多項分布に応用

$$h(x) \qquad \eta_k = \log\left(\mu_k \Big/ 1-\sum_{k=1}^{K-1}\mu_k\right) \qquad x \qquad a(\eta)$$

$$p(x\,|\,\mu) = \begin{pmatrix} N \\ x_1 \cdots x_K \end{pmatrix} \prod_{k=1}^{K} \mu_k^{\,x_k} = \begin{pmatrix} N \\ x_1 \cdots x_K \end{pmatrix} \exp\left\{ \sum_{k=1}^{K-1} \log\left(\frac{\mu_k}{1-\sum_{k=1}^{K-1}\mu_k} \right) x_k + N\left(\log\left(1-\sum_{k=1}^{K-1}\mu_k\right)\right) \right\}$$

$$e^{\eta_k} = \frac{\mu_k}{1-\sum_{k=1}^{K-1}\mu_k} \qquad \sum_{k=1}^{K} e^{\eta_k} = \frac{\sum_{k=1}^{K}\mu_k}{1-\sum_{k=1}^{K-1}\mu_k} = \frac{1}{1-\sum_{k=1}^{K-1}\mu_k} \qquad a(\eta_k) = -N\log\left(1-\sum_{k=1}^{K-1}\mu_k\right) = -N\log\frac{1}{\sum_{k=1}^{K} e^{\eta_k}} = N\log\sum_{k=1}^{K} e^{\eta_k}$$

$$\boxed{\frac{\partial a(\eta_k)}{\partial \eta_k}} = N\frac{e^{\eta_k}}{\sum_{k=1}^{K} e^{\eta_k}} = N\frac{\frac{\mu_k}{1-\sum_{k=1}^{K-1}\mu_k}}{\sum_{k=1}^{K} \frac{\mu_k}{1-\sum_{k=1}^{K-1}\mu_k}{}^{\eta_k}} = N\frac{\mu_k}{\sum_{k=1}^{K}\mu_k} = \boxed{N\mu_k}$$

$$\boxed{\frac{\partial^2 a(\eta_k)}{\partial \eta_k^{\ 2}}} = \frac{\partial N \dfrac{e^{\eta_k}}{\sum_{k=1}^{K} e^{\eta_k}}}{\partial \eta_k} = N\frac{e^{\eta_k}\left(\sum_{k=1}^{K} e^{\eta_k}\right) - e^{\eta_k} e^{\eta_k}}{\left(\sum_{k=1}^{K} e^{\eta_k}\right)^2} = N\frac{e^{\eta_k}\left(\sum_{k=1}^{K} e^{\eta_k} - e^{\eta_k}\right)}{\left(\sum_{k=1}^{K} e^{\eta_k}\right)^2} = N\frac{\frac{\mu_k}{1-\sum_{k=1}^{K-1}\mu_k}\left(\frac{1-\mu_k}{1-\sum_{k=1}^{K-1}\mu_k}\right)}{\left(1-\sum_{k=1}^{K-1}\mu_k\right)^{-2}} = \boxed{N\mu_k(1-\mu_k)}$$

➢2項分布の場合はおなじみの公式

$$p(x\mid\eta)=\binom{N}{x}\mu^x(1-\mu)^{N-x}=\binom{N}{2}\exp\left\{x\log\frac{\mu}{1-\mu}+N(\log(1-\mu))\right\}$$

$$h(x)\qquad\qquad \eta=\log\left(\frac{\mu}{1-\mu}\right)\qquad\qquad a(\eta)$$

$$e^{\eta}=\frac{\mu}{1-\mu}\qquad e^{\eta}(1-\mu)=\mu\qquad e^{\eta}=\mu\left(1+e^{\eta}\right)\qquad \mu=\frac{e^{\eta}}{1+e^{\eta}}$$

$$a(\eta)=-N\log(1-\mu)=-N\log\left(\frac{1}{1+e^{\eta}}\right)=N\log\left(1+e^{\eta}\right)$$

$$\boxed{E_{\eta}[x]}=\frac{\partial a(\eta)}{\partial\eta}=N\frac{e^{\eta}}{1+e^{\eta}}\boxed{=N\mu}\qquad \boxed{V_{\eta}[x]}=\frac{\partial^2 a(\eta)}{\partial\eta^2}=\frac{\partial}{\partial\eta}N\frac{e^{\eta}}{1+e^{\eta}}=N\frac{e^{\eta}}{\left(1+e^{\eta}\right)^2}=\boxed{N\mu(1-\mu)}$$

Exponential familyとベイズ統計：共役分布と事後分布

ハイパーパラメター：$\lambda=\left(\lambda_1{}^T,\lambda_2\right)$　によって共役事前分布を定義する

$$p(\eta\mid\lambda)=h(\eta)\exp\left\{\lambda_1{}^T\eta+\lambda_2\left(-a(\eta)\right)-a(\lambda)\right\}$$

$$a(\lambda)=\log\left(\int h(\eta)\exp\left\{\left(\lambda_1{}^T,\lambda_2\right)\begin{pmatrix}\eta\\-a(\eta)\end{pmatrix}\right\}\mathrm{d}\eta\right)\qquad(EB22)$$

さて、$p(x\mid\eta)$に沿うK個のiidの観測データ$x_1\cdots x_K$が得られたときのηの事後分布は

$$p(\eta\mid x_1,\cdots,x_N,\lambda)\propto p(\eta\mid\lambda)\prod_{i=1}^{K}p\left(x_i\mid\eta\right)$$

$$\propto h(\eta)\exp\left\{\lambda_1{}^T\eta+\lambda_2\left(-a(\eta)\right)-a(\lambda)\right\}\prod_{i=1}^{K}\exp\left\{\eta^T u(x_i)-a(\eta)\right\}$$

$$\propto h(\eta)\exp\left\{\left(\boxed{\lambda_1+\sum_{i=1}^{K}u(x_i)}\right)^T\eta+\boxed{\left(\lambda_2+K\right.}\left.\right)\left(-a(\eta)\right)-a(\lambda)\right\}$$

赤枠の中は事後パラメター

仮想的な観測データ

実際の観測データ

仮想的な観測回数:1

実際の観測回数

1変数正規分布の期待値に適用した例　その1

$$\text{事前分布}: p(\mu\,|\,\lambda) = N(\mu\,|\,\mu_0, \sigma_0^2) \qquad \lambda = (\lambda_1, \lambda_2)^T, \lambda_1 = \frac{\mu_0}{{\sigma_0}^2}, \lambda_2 = \frac{1}{2{\sigma_0}^2}$$

$$p(\mu\,|\,\lambda) = N(\mu\,|\,\mu_0, \sigma_0^2) \propto \exp\{\lambda_1\ \mu - \lambda_2 a(\mu) - a(\lambda)\} = \exp\left\{\frac{\mu_0}{{\sigma_0}^2}\mu - \frac{\mu^2}{2{\sigma_0}^2} - \frac{{\mu_0}^2}{2{\sigma_0}^2}\right\}$$

K個の観測データが得られた場合の尤度$: p(X\,|\,\mu)$

$$= \prod_{i=1}^{K} p(x_i\,|\,\mu) = \frac{1}{(2\pi\sigma^2)^{K/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{K}(x_i - \mu)^2\right\}$$

ただしσは既知

$$p(x\,|\,\mu, \sigma^2) == \frac{1}{\sqrt{2\pi}} \exp\left\{\underbrace{\left(\frac{\mu}{\sigma^2}, \frac{1}{\sigma^2}\right)}_{\eta^T}\begin{pmatrix} x \\ -\dfrac{x^2}{2} \end{pmatrix} - \frac{1}{2}\underbrace{\left(\frac{\mu^2}{\sigma^2} + \log\frac{1}{\sigma^2}\right)}_{a(\eta)}\right\}$$

$\mu_0, {\sigma_0}^2, \sigma^2$は既知とする。

1変数正規分布の期待値に適用した例　その2

$$p(\mu \mid x_1, \cdots, x_N, \lambda) \propto p(\mu \mid \lambda) \prod_{i=1}^{K} p(x_i \mid \mu)$$

$$\propto h(\eta) \exp\{\lambda_1 \mu - \lambda_2 a(\mu)\} \prod_{i=1}^{K} \exp\{\eta^{\mathrm{T}} u(x_i) - a(\eta)\}$$

$$\eta^T = \left(\frac{\mu}{\sigma^2}, \quad \frac{1}{\sigma^2} \right)$$

$$\propto h(\eta) \exp\left\{ \lambda_1 \mu + \left(\sum_{i=1}^{K} x_i, \sum_{i=1}^{K} \frac{x_i^{\ 2}}{2} \right) \begin{pmatrix} \dfrac{\mu}{\sigma^2} \\ \dfrac{1}{\sigma^2} \end{pmatrix} - \lambda_2 a(\lambda) - K a(\eta) \right\}$$

$$\propto h(\eta) \exp\left\{ \mu \frac{\mu_0}{\sigma_0^{\ 2}} + \mu \frac{1}{\sigma^2} \sum_{i=1}^{K} x_i - \frac{\mu^2}{2\sigma_0^{\ 2}} - \frac{K}{2} \left(\frac{\mu^2}{\sigma^2} + \log \frac{1}{\sigma^2} \right) \right\}$$

$$\propto \exp\left\{ \mu \left(\frac{\mu_0}{\sigma_0^{\ 2}} + \frac{1}{\sigma^2} \sum_{i=1}^{K} x_i \right) - \frac{\mu^2}{2} \left(\frac{1}{\sigma_0^{\ 2}} + \frac{K}{\sigma^2} \right) \right\}$$

←前に求めた $\boxed{\text{N10}}$ に一致

Exponential family別表現とベイズ統計の続き：予測分布

K個の観測データも考慮したハイパーパラメター

$$\boxed{\hat{\lambda}_1 = \lambda_1 + \sum_{i=1}^{K} u(x_i) \qquad (EB21)} \qquad \boxed{\hat{\lambda}_2 = \lambda_2 + K \qquad (EB22)}$$

ハイパーパラメターが与えられたときの事後分布は次式のようにηでmarginaliz e

$$p(x \mid \lambda) = \int p(x \mid \eta) p(\eta \mid \lambda) \mathrm{d}\eta$$

$$= h(x) \int \exp\left\{\eta^T u(x) - a(\eta)\right\} h(\eta) \exp\left\{\lambda_1{}^T \eta + \lambda_2 (-a(\eta)) - a(\lambda)\right\} \mathrm{d}\eta$$

$$\boxed{= h(x) \int h(\eta) \exp\left\{(\lambda_1 + u(x))^T \eta + (\lambda_2 + 1)(-a(\eta))\right\} \mathrm{d}\eta \exp\{-a(\lambda)\} \qquad (EB23)}$$

$\Rightarrow$　　ハイパーパラメターλと

K個のiidの観測データ$x_1 \cdots x_K$が得られたときの

新規（あるいは未知）の　xの予測分布は$(EB23)$において

λ_1を$\hat{\lambda}_1 = \lambda_1 + \sum_{i=1}^{K} u(x_i)$で

λ_2を$\hat{\lambda}_2 = \lambda_2 + K$で

置き換えれば得られる。

ベイズ統計による事前、事後、予測分布の例：多変数ガウス分布
難しいので省略する予定

精度行列（分散の逆行列）Λが既知のd次元ガウス分布をexponential family で表現

$$p(x\,|\,\eta)=\frac{1}{2}\left(2\pi\,|\,\Sigma\,|\right)^{-\frac{d}{2}}\exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}\qquad \Lambda=\Sigma^{-1}\quad \text{とおくと}$$

$$=\exp\left\{-\frac{1}{2}\left(d\log 2\pi+d\log|\Lambda|+(x-\mu)^T\Lambda(x-\mu)\right)\right\}$$

$$=\exp\left\{-\frac{1}{2}\left(d\log 2\pi+d\log|\Lambda|+x^T\Lambda x-x^T\Lambda\mu-\mu^T\Lambda x+\mu^T\Lambda\mu\right)\right\}$$

natural parameter : ηがまだ決めていなかった!

$$\Rightarrow\quad \eta=\Lambda\mu\qquad \text{以下も注意}\quad \Lambda^{\mathrm{T}}=\Lambda\qquad \left(\Lambda^{-1}\right)^{\mathrm{T}}=\Lambda^{-1}\qquad \mu=\Lambda^{-1}$$

$$p(x\,|\,\eta)=\exp\Bigg\{\underbrace{-\frac{1}{2}\left(d\log 2\pi+d\log|\Lambda|+x^T\Lambda x\right)}_{h(x)}\Bigg\}\exp\Bigg\{\eta^T\underbrace{x}_{u(x)=(x_1,..,x_d)^T}\underbrace{-\frac{1}{2}\eta^T\Lambda^{-1}\eta}_{a(\eta)}\Bigg\}$$

事前分布のパラメーターλから予測分布$p(x \mid \lambda)$を求める

$$p(\eta \mid \lambda) \propto \exp\left\{\lambda_1{}^T \eta + \lambda_2 a(\eta) - a(\lambda)\right\} = \exp\left\{\lambda_1{}^T \eta - \lambda_2 \frac{1}{2} \eta^T \Lambda^{-1} \eta - a(\lambda)\right\}$$

$$= \exp\left\{-\frac{1}{2}\left(\eta^T \lambda_2 \Lambda^{-1} \eta - 2\lambda_1{}^T \eta + 2a(\lambda)\right)\right\}$$ 事前分布もガウス分布として

$$= \exp\left\{-\frac{1}{2}\left(\left(\eta - \frac{\Lambda\lambda_1}{\lambda_2}\right)^T \lambda_2 \Lambda^{-1}\left(\eta - \frac{\Lambda\lambda_1}{\lambda_2}\right)\right)\right\}$$ とおくと

$$\propto \exp\left\{\frac{1}{2}\log|\lambda_2 \Lambda^{-1}| - \frac{1}{2}\left(\eta^T \lambda_2 \Lambda^{-1} \eta - 2\lambda_1{}^T \eta + \frac{\lambda_1{}^T \Lambda \lambda_1}{\lambda_2}\right)\right\}$$

$$\Rightarrow \quad a(\lambda) = -\frac{1}{2}\left(d \log \lambda_2 + \log|\Lambda^{-1}|\right) + \frac{1}{2}\frac{\lambda_1{}^T \Lambda \lambda_1}{\lambda_2}$$

$\because$ d次元の行列Λ^{-1}に対しては、$|\lambda_2 \Lambda^{-1}| = \lambda_2{}^d |\Lambda^{-1}|$

$$p(\eta \mid \lambda) = h(\eta)\exp\left\{\lambda_1{}^T\eta + \lambda_2\left(-a(\eta)\right) - a(\lambda)\right\}$$

$$a(\lambda) = \log\left(\int h(\eta)\exp\left\{\left(\lambda_1{}^T, \lambda_2\right)\begin{pmatrix}\eta \\ -a(\eta)\end{pmatrix}\right\}\mathrm{d}\eta\right) \qquad (EB22)$$ および

$(EB3)(EB4)$よりλが与えられたときのηの十分統計量が以下のように求まる。

$$\boxed{E[\eta] = \frac{\partial a(\lambda)}{\partial \lambda_1} = \frac{\left(\Lambda + \Lambda^T\right)\lambda_1}{2\lambda_2} \qquad (EB35) \qquad E[-a(\eta)] = \frac{\partial a(\lambda)}{\partial \lambda_2} = \frac{d}{2\lambda_2} - \frac{\lambda_1{}^T\Lambda\lambda_1}{\lambda_2{}^2} \qquad (EB36)}$$