# 2.3.3 ガウス変数に対するベイズの定理

6月20日

鈴木直也

# 線形ガウスモデル

2.3.1節、2.3.2節においてベクトルxを2つの部分ベクトルに分割したガウス分布を考察してから、条件付分布と周辺分布を求めた。[*1]

ここで、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 $x$ とは独立であるようなガウス条件付分布$p(y|x)$が与えられたとき、周辺分布$p(y)$と条件付分布 $p(x|y)$を求める問題を考える。

---

注1：ここで条件付分布の平均はxbの線形関数である。

# 線形ガウスモデル2

周辺分布と条件付分布をそれぞれ

$$p(x) = N(x \mid \mu, \Lambda^{-1}) \qquad (2.99)$$

$$p(y \mid x) = N(y \mid \mathrm{A}x + b, \mathrm{L}^{-1}) \qquad (2.100)$$

とする。ただし、$\mu, \mathrm{A}, b$ は平均に関係したパラメータで $\Lambda, \mathrm{L}$ は精度行列である

$x$ が $M$ 次元で $y$ が $D$ 次元なら

行列Aの大きさは $D \times M$ となる。

また行列Lの大きさは $D \times D$

# 同時分布の表現

$x$ と $y$ 上の同時分布の表現を求めるため次のように定義する。

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \qquad (2.101)$$

そして同時分布の対数を考える。

$$\begin{aligned} \ln p(z) &= \ln p(x) + \ln p(y \mid x) \\ &= -\frac{1}{2}(x-\mu)^{\mathrm{T}} \Lambda (x-\mu) - \frac{1}{2}(y - \mathrm{A}x - b)^{\mathrm{T}} \mathrm{L} (y - \mathrm{A}x - b) + const^{*1} \end{aligned} \qquad (2.102)$$

---

注1：このときconstはxやyとは独立の定数である。

# z上のガウス分布の精度行列

このガウス分布の精度行列を求めるために前式について考察する。
まず変形すると

$$-\frac{1}{2}(x-\mu)^{\mathrm{T}}\Lambda(x-\mu)-\frac{1}{2}(y-\mathrm{A}x-b)^{\mathrm{T}}\mathrm{L}(y-\mathrm{A}x-b)+const$$

$$=-\frac{1}{2}(x^{\mathrm{T}}\Lambda x-x^{\mathrm{T}}\Lambda\mu-x\Lambda\mu^{\mathrm{T}}+\mu^{\mathrm{T}}\Lambda\mu)$$

$$-\frac{1}{2}(y^{\mathrm{T}}\mathrm{L}y-y^{\mathrm{T}}\mathrm{LA}x-y^{\mathrm{T}}\mathrm{L}b-x^{\mathrm{T}}\mathrm{A}^{\mathrm{T}}\mathrm{L}y$$

$$+x^{\mathrm{T}}\mathrm{A}^{\mathrm{T}}\mathrm{LA}x+x^{\mathrm{T}}\mathrm{A}^{\mathrm{T}}\mathrm{L}b-b^{\mathrm{T}}\mathrm{L}y+b^{\mathrm{T}}\mathrm{LA}x+b^{\mathrm{T}}\mathrm{L}b)+const$$

# z上のガウス分布の精度行列(2)

この式の２次の項について注目すると

$$
-\frac{1}{2}x^{\mathrm{T}}(\Lambda+\mathrm{A}^{\mathrm{T}}\mathrm{LA})x-\frac{1}{2}y^{\mathrm{T}}\mathrm{L}y+\frac{1}{2}y^{\mathrm{T}}\mathrm{LA}x+\frac{1}{2}x^{\mathrm{T}}\mathrm{A}^{\mathrm{T}}\mathrm{L}y
$$

$$
=-\frac{1}{2}\begin{pmatrix} x \\ y \end{pmatrix}^{\mathrm{T}}\begin{pmatrix} \Lambda+\mathrm{A}^{\mathrm{T}}\mathrm{LA} & -\mathrm{A}^{\mathrm{T}}\mathrm{L} \\ -\mathrm{LA} & \mathrm{L} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}=-\frac{1}{2}z^{\mathrm{T}}\mathrm{R}z \qquad (2.103)
$$

よってz上のガウス分布の精度行列は

$$
R=\begin{pmatrix} \Lambda+\mathrm{A}^{\mathrm{T}}\mathrm{LA} & -\mathrm{A}^{\mathrm{T}}\mathrm{L} \\ -\mathrm{LA} & \mathrm{L} \end{pmatrix} \qquad (2.104)
$$

となる。

# z上のガウス分布の共分散行列

共分散行列は行列の逆行列に関する公式(2.76)を
利用して精度行列の逆行列を求めることで得られる。

$$\mathrm{cov}[z] = R^{-1} = \begin{pmatrix} \Lambda + \mathrm{A}^{\mathrm{T}}\mathrm{LA} & \text{-}\mathrm{A}^{\mathrm{T}}\mathrm{L} \\ \text{-}\mathrm{LA} & \mathrm{L} \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} M & \text{-}M(\text{-}\mathrm{A}^{\mathrm{T}}\mathrm{L})\mathrm{L}^{-1} \\ \text{-}\mathrm{L}^{-1}(\text{-}\mathrm{LA})M & \mathrm{L}^{-1} + \mathrm{L}^{-1}(\text{-}\mathrm{LA})M(\text{-}\mathrm{A}^{\mathrm{T}}\mathrm{L})\mathrm{L}^{-1} \end{pmatrix}$$

このとき

$$M = [\Lambda + \mathrm{A}^{\mathrm{T}}\mathrm{LA}\text{-}(\text{-}\mathrm{A}^{\mathrm{T}}\mathrm{L})\mathrm{L}^{-1}(\text{-}\mathrm{LA})]^{-1} = \Lambda^{-1}$$

のように定義される。

# z上のガウス分布の共分散行列(2)

よって

$$\mathrm{cov}[z] = \begin{pmatrix} \Lambda^{-1} & \Lambda^{-1}A^{T} \\ A\Lambda^{-1} & L^{-1} + A\Lambda^{-1}A^{T} \end{pmatrix} \qquad (2.105)$$

となる。

# z上のガウス分布の平均

同様に平均は(2.102)の線形の項を調べることで

$$x^{\mathrm{T}}\Lambda\mu - x^{\mathrm{T}}\mathrm{A}^{\mathrm{T}}\mathrm{L}b + y^{\mathrm{T}}\mathrm{L}b = \begin{pmatrix} x \\ y \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \Lambda\mu - \mathrm{A}^{\mathrm{T}}\mathrm{L}b \\ \mathrm{L}b \end{pmatrix} \qquad (2.106)$$

で与えられる。また、(2.71)によりzの平均は

$$E[z] = R^{-1} \begin{pmatrix} \Lambda\mu - \mathrm{A}^{\mathrm{T}}\mathrm{L}b \\ \mathrm{L}b \end{pmatrix} \qquad (2.\ 107)$$

# z上のガウス分布の平均(2)

これに(2.105)を利用すると

$$E[z]=\begin{pmatrix}\Lambda^{-1} & \Lambda^{-1}\mathrm{A}^{\mathrm{T}}\\ \mathrm{A}\Lambda^{-1} & \mathrm{L}^{-1}+\mathrm{A}\Lambda^{-1}\mathrm{A}^{\mathrm{T}}\end{pmatrix}\begin{pmatrix}\Lambda\mu-\mathrm{A}^{\mathrm{T}}\mathrm{L}b\\ \mathrm{L}b\end{pmatrix}$$

$$=\begin{pmatrix}\Lambda^{-1}(\Lambda\mu-\mathrm{A}^{\mathrm{T}}\mathrm{L}b)+\Lambda^{-1}\mathrm{A}^{\mathrm{T}}\mathrm{L}b\\ \mathrm{A}\Lambda^{-1}(\Lambda\mu-\mathrm{A}^{\mathrm{T}}\mathrm{L}b)+(\mathrm{L}^{-1}+\mathrm{A}\Lambda^{-1}\mathrm{A}^{\mathrm{T}})\mathrm{L}b\end{pmatrix}$$

$$=\begin{pmatrix}\mu\\ \mathrm{A}\mu+b\end{pmatrix} \qquad (2.108)$$

を得る。

# xの周辺分布

次にxを周辺化した周辺分布 $p(y)$ を求める。

(2.92)、(2.93)より周辺分布$p(y)$の平均と共分散は

$$E[y] = \mathrm{A}\mu + b \qquad (2.109)$$

$$\mathrm{cov}[y] = \mathrm{L}^{-1} + \mathrm{A}\Lambda^{-1}\mathrm{A}^{\mathrm{T}} \qquad (2.110)$$

# 条件付き分布

(2.73)(2.75)を(2.105)と(2.108)に利用すると

$$
\begin{aligned}
E[x \mid y] &= \mu - (\Lambda + \mathrm{A}^{\mathrm{T}}\mathrm{LA})^{-1}(-\mathrm{A}^{\mathrm{T}}\mathrm{L})(y - \mathrm{A}\mu - b) \\
&= (\Lambda + \mathrm{A}^{\mathrm{T}}\mathrm{LA})^{-1}(\Lambda + \mathrm{A}^{\mathrm{T}}\mathrm{LA})\mu - (\Lambda + \mathrm{A}^{\mathrm{T}}\mathrm{LA})^{-1}(-\mathrm{A}^{\mathrm{T}}\mathrm{Ly} + \mathrm{A}^{\mathrm{T}}\mathrm{LA}\mu + \mathrm{A}^{\mathrm{T}}\mathrm{L}b) \\
&= (\Lambda + \mathrm{A}^{\mathrm{T}}\mathrm{LA})^{-1}\left(\mathrm{A}^{\mathrm{T}}\mathrm{Ly} - \mathrm{A}^{\mathrm{T}}\mathrm{LA}\mu - \mathrm{A}^{\mathrm{T}}\mathrm{L}b + \Lambda\mu + \mathrm{A}^{\mathrm{T}}\mathrm{LA}\mu\right) \\
&= (\Lambda + \mathrm{A}^{\mathrm{T}}\mathrm{LA})^{-1}\left[\mathrm{A}^{\mathrm{T}}\mathrm{L}\left(\mathrm{y} - b\right) + \Lambda\mu\right] \qquad (2.111)
\end{aligned}
$$

$$
\mathrm{cov}[x \mid y] = (\Lambda + \mathrm{A}^{\mathrm{T}}\mathrm{LA})^{-1} \qquad (2.112)
$$

で与えられる。

# まとめ

xの周辺ガウス分布と、xが与えられたときのyの条件付きガウス分布が次式で与えられたとする

$$p(x) = \mathcal{N}(x \mid \mu, \Lambda^{-1}) \qquad (2.113)$$

$$p(y \mid x) = \mathcal{N}(y \mid \mathrm{A}x + b, \mathrm{L}^{-1}) \qquad (2.114)$$

yの周辺分布と、yが与えられたときのxの条件付分布は

$$p(y) = \mathcal{N}(y \mid \mathrm{A}\mu + b, \mathrm{L}^{-1} + \mathrm{A}\Lambda^{-1}\mathrm{A}^{\mathrm{T}}) \qquad (2.115)$$

$$p(x \mid y) = \mathcal{N}(x \mid \Sigma\left\{\mathrm{A}^{\mathrm{T}}\mathrm{L}(\mathrm{y}\text{-}\mathrm{b}) + \Lambda\mu\right\}, \Sigma) \qquad (2.116)$$

で与えられる。ただし、

$$\Sigma = (\Lambda + \mathrm{A}^{\mathrm{T}}\mathrm{L}\mathrm{A})^{-1} \qquad (2.117)$$