

Math Basis 03

Chen Gong

20 October 2019

本节的主要内容是在高斯分布中，已知联合概率密度求条件概率密度和边缘概率密度。还有已知 x 的边缘概率和在条件为 x 下 y 的条件概率下，求 y 的边缘概率和在条件为 y 下 x 的条件概率。用数学的语言描述即为，如下所示。

对于多变量的高斯分布 $X \sim \mathcal{N}(\mu, \Sigma)$ ，概率密度函数为：

$$p(X) = \frac{1}{\sqrt{2\pi}^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (X - \mu)^T \Sigma^{-1} (X - \mu) \right\} \quad (1)$$

其中， $X \in \mathbb{R}^p$ ，

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}_{p \times p} \quad (2)$$

已知联合概率密度求条件概率密度和边缘概率密度，可描述为已知：

$$X = \begin{pmatrix} x_a \\ x_b \end{pmatrix} \quad m + n = p \quad \mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} \quad (3)$$

求： $p(x_a)$ 和 $p(x_b|x_a)$

另一个可描述为，已知： $p(x) = \mathcal{N}(x|\mu, \Lambda^{-1})$ ， $p(y|x) = \mathcal{N}(Ax + b, L^{-1})$ ，求 $p(y)$ 和 $p(x|y)$ 。

1 已知联合概率密度求条件概率密度和边缘概率密度

在进行此次推导之前，首先需要引入一个推论。已知：

$$X \sim \mathcal{N}(\mu, \Sigma) \quad (4)$$

$$Y = AX + b \quad (5)$$

那么可以解得： $Y \sim \mathcal{N}(A\mu + b, A\Sigma A^T)$ 。

1.1 求解边缘概率密度 $p(x_a)$

$$x_a = (I_m, 0) \begin{pmatrix} x_a \\ x_b \end{pmatrix} \quad (6)$$

很显然，等式的第一部分就是 A ，等式的第二部分就是 X ，那么

$$\mathbb{E}[x_a] = (I_m, 0) \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} = \mu_a \quad \text{Var}[x_a] = (I_m, 0) \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} \begin{pmatrix} I_m \\ 0 \end{pmatrix} = \Sigma_{aa} \quad (7)$$

所以， $p(x_a) \sim \mathcal{N}(\mu_a, \Sigma_{aa})$ 。

1.2 求解条件概率密度 $p(x_b|x_a)$

这里使用的是构造证明法，有一点猜出了结论，再使用结论来证结论的感觉。在这里引入了一个构造函数 $x_{b \cdot a} = x_b - \Sigma_{ba}\Sigma_{aa}^{-1}x_a$ 。那么根据我们的推论，我们可以很简单的得出以下的构造方法

$$\mu_{b \cdot a} = \mu_b - \Sigma_{ba}\Sigma_{aa}^{-1}\mu_a \quad (8)$$

$$\Sigma_{bb \cdot a} = \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab} \quad (9)$$

其中 $\Sigma_{aa \cdot b}$ 被称为 Σ_{aa} 的 Schur Complementary。这是什么鬼？我也不知道，只知道这个概念来自矩阵分析，有兴趣的朋友可以进一步的去学习。我们接下来可以很简单的将 $x_{b \cdot a}$ 化简成 $Ax + b$ 的性质。

$$x_{b \cdot a} = (-\Sigma_{ba}\Sigma_{aa}^{-1} \quad I_m) \begin{pmatrix} x_a \\ x_b \end{pmatrix} \quad (10)$$

那么我们就进行以下的推导，

$$\mathbb{E}[x_{b \cdot a}] = (-\Sigma_{ba}\Sigma_{aa}^{-1} \quad I_m) \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} = \mu_b - \Sigma_{ba}\Sigma_{aa}^{-1}\mu_a \quad (11)$$

$$\text{Var}[x_{b \cdot a}] = (-\Sigma_{ba}\Sigma_{aa}^{-1} \quad I_m) \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} \begin{pmatrix} -\Sigma_{aa}^{-1T}\Sigma_{ba}^T \\ I_m \end{pmatrix} \quad (12)$$

计算得出， $\text{Var}[x_{b \cdot a}] = \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}$ 。所以， $p(x_{b \cdot a}) = \mathcal{N}(\mu_{b \cdot a}, \Sigma_{bb \cdot a})$ 。

又因为 $x_b = x_{b \cdot a} + \Sigma_{ba}\Sigma_{aa}^{-1}x_a$ ， x_b 是一个关于 x_a 的线性表达。那么后续的求解过程将变得非常的简单了。

$$\mathbb{E}[x_b|x_a] = \mathbb{E}[x_{b \cdot a}] + \Sigma_{ba}\Sigma_{aa}^{-1}x_a = \mu_b + \Sigma_{ba}\Sigma_{aa}^{-1}(x_a - \mu_a) \quad (13)$$

而 $\text{Var}[x_b|x_a]$ 中

$$\text{Var}[x_b|x_a] = \Sigma_{bb \cdot a} = \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab} \quad (14)$$

于是综上所述， $p(x_b|x_a) \sim \mathcal{N}(\mu_b + \Sigma_{ba}\Sigma_{aa}^{-1}(x_a - \mu_a), \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})$

2 边缘高斯和条件高斯

问题可以被描述为， $p(x) = \mathcal{N}(x|\mu, \Lambda^{-1})$ ， $p(y|x) = \mathcal{N}(Ax + b, L^{-1})$ ，求 $p(y)$ 和 $p(x|y)$ 。

这个高斯分布的公式的推导非常的重要，特别是在 Linear Gaussian Model 中被大量的运用。同时也在贝叶斯公式中有着大量的运用，所以某种意义上说，这个公式的运用比上一个公式更加的重要。

2.1 边缘高斯

根据已知条件, 我们可以设 $y = Ax + b + \epsilon$, 其中 $\epsilon \sim \mathcal{N}(0, L^{-1})$ 。

$$\mathbb{E}[y] = \mathbb{E}[Ax + b + \epsilon] = \mathbb{E}[Ax + b] + \mathbb{E}[\epsilon] = A\mu + b \quad (15)$$

$$\text{Var}[y] = \text{Var}[Ax + b + \epsilon] = \text{Var}[Ax] + \text{Var}[\epsilon] = A\Lambda^{-1}A^T + L^{-1} \quad (16)$$

所以综上所述, $p(y) \sim \mathcal{N}(A\mu + b, A\Lambda^{-1}A^T + L^{-1})$ 。

2.2 条件高斯

根据上述的推导我们已经知道的条件有, $p(y)$, $p(x)$, $p(y|x)$ 那么我们可以如何求得 $p(x|y)$ 呢? 很显然我们还差一个 x 和 y 的联合分布。如果知道 x 和 y 的联合分布, 那么我们就和上一节已知联合概率密度求条件概率密度和边缘概率密度的内容接起来了。

设

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu \\ A\mu + b \end{pmatrix}, \begin{pmatrix} \Lambda^{-1} & \Delta \\ \Delta & A\Lambda^{-1}A^T + L^{-1} \end{pmatrix} \right) \quad (17)$$

其中 Δ 是多少, 我们并不知道, 需要想办法来求解。这里的 Δ 是 x 和 y 的协方差矩阵, 我们可以直接利用协方差的定义来进行求解。

$$\begin{aligned} \Delta &= \text{Cov}(x, y) = \mathbb{E}[(x - \mathbb{E}[x])(y - \mathbb{E}[y])^T] \\ &= \mathbb{E}[(x - \mu)(y - (A\mu + b))^T] \\ &= \mathbb{E}[(x - \mu)(Ax + \epsilon - A\mu)^T] \\ &= \mathbb{E}[(x - \mu)(Ax - A\mu + \epsilon)^T] \\ &= \mathbb{E}[(x - \mu)(Ax - A\mu)^T + ((x - \mu)\epsilon)^T] \\ &= \mathbb{E}[(x - \mu)(Ax - A\mu)^T] + \mathbb{E}[(x - \mu)\epsilon^T] \end{aligned} \quad (18)$$

显然, $x - \mu$ 和 ϵ 之间是相互独立的。那么, $\mathbb{E}[(x - \mu)\epsilon^T] = \mathbb{E}[(x - \mu)]\mathbb{E}[\epsilon^T] = 0$ 。那么, 推导继续,

$$\begin{aligned} \Delta &= \mathbb{E}[(x - \mu)(Ax - A\mu)^T] \\ &= \mathbb{E}[(x - \mu)(x - \mu)^T] \cdot A^T \\ &= \text{Var}[x] \cdot A^T \\ &= \Lambda^{-1}A^T \end{aligned} \quad (19)$$

所以, x 和 y 之间的联合概率分布可表达为,

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu \\ A\mu + b \end{pmatrix}, \begin{pmatrix} \Lambda^{-1} & \Lambda^{-1}A^T \\ \Lambda^{-1}A^T & A\Lambda^{-1}A^T + L^{-1} \end{pmatrix} \right) \quad (20)$$

通过上述的推导, 我们可以成功的得到 x 和 y 的联合概率密度分布。那么利用上节得出的公式, 就可以成功的推导出 $p(x|y)$ 。代入上述公式中, 描述的结果如下:

$$p(x|y) = \mathcal{N}(\Sigma\{A^T L(y - b) + \Lambda\mu\}, \Sigma) \quad (21)$$

其中,

$$\Sigma = (\Lambda + A^T L A)^{-1} \quad (22)$$