● 一维高斯分布

■ 一维高斯分布

➢ 若一维随机变量满足高斯分布，即$X\sim\mathcal{N}(\mu,\ \sigma^2)$，则

➢ 概率密度函数

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

➢ 概率密度积分

$$\int_{-\infty}^{+\infty}f(x)dx=1$$

■ 一维标准正态分布

➢ 仿射变换$Z=(X-\mu)/\sigma$

➢ 概率密度函数

$$f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$$

➢ 标准正态分布$Z\sim\mathcal{N}(0,\ 1)$

推导

已知仿射变换函数$z=(x-\mu)/\sigma$，其逆函数$x(z)=\sigma z+\mu$,于是，概率密度函数变为

$$f\big(x(z)\big)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{z^2}{2}}$$

概率密度积分变为

$$1=\int_{-\infty}^{+\infty}f\big(x(z)\big)dx=\int_{-\infty}^{+\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{z^2}{2}}dx$$

$$=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz=\int_{-\infty}^{+\infty}f(z)dz$$

- **(解耦)多维高斯分布**

■ 多维正态分布

➢ 随机向量$\boldsymbol{Z} = [Z_1, Z_2, \ldots, Z_n]^T$，其中$Z_i \sim \mathcal{N}(0,\ 1)$，且$Z_i$和$Z_j(i, j = 1,2, \ldots, n,\ i \neq j)$相互独立

➢ 随机向量$\boldsymbol{Z}$的联合概率密度函数

$$f(z_1, z_2, \ldots, z_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}} = \frac{1}{\left(\sqrt{2\pi}\right)^n} e^{-\frac{1}{2}\boldsymbol{Z}^T\boldsymbol{Z}}$$

➢ 概率密度积分

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} f(z_1, z_2, \ldots, z_n) dz_n dz_{n-1} \ldots dz_1 = 1$$

■ 矩阵表示

➢ 随机变量$\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0},\ \boldsymbol{I})$

➢ 均值(期望)向量：n维独立变量可以排列成列向量

$$\boldsymbol{\mu} = \boldsymbol{0}$$

➢ 协方差矩阵：n维对角线矩阵

$$\boldsymbol{\Sigma} = \mathbb{E}[(\boldsymbol{Z} - \boldsymbol{\mu})(\boldsymbol{Z} - \boldsymbol{\mu})^T] = \boldsymbol{I}$$，也可以记作$cov(\boldsymbol{Z},\ \boldsymbol{Z})$

- **(一般)多维高斯分布**
  - 高斯分布与标准正态分布
    - 一维：$Z\sim\mathcal{N}(0,\ 1)$，仿射变换$Z=(X-\mu)/\sigma$，则$X\sim\mathcal{N}(\mu,\ \sigma^2)$
    - 多维：$\boldsymbol{Z}\sim\mathcal{N}(\boldsymbol{0},\ \boldsymbol{I})$，仿射变换$\boldsymbol{Z}=\boldsymbol{A}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})$，则$\boldsymbol{X}\sim\mathcal{N}(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$
  - 一般多维高斯分布
    - 均值(期望)向量：变换中的偏置项$\boldsymbol{\mu}$
    - 协方差矩阵：$\boldsymbol{\Sigma}=\boldsymbol{A}\boldsymbol{A}^T$

仿射变换

$$\boldsymbol{T}=\begin{bmatrix}\boldsymbol{A} & \boldsymbol{t}\\ 0 & 1\end{bmatrix}$$

其中:

① $\boldsymbol{A}$是线性变换矩阵(可逆，不要求正交，正交则为旋转变换);

② $\boldsymbol{t}$是平移变换向量;

③ $\boldsymbol{T}$是n+1维矩阵(增广形式)

推导

已知仿射变换函数$\boldsymbol{Z}=\boldsymbol{A}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})$，概率密度函数$f\big(\boldsymbol{Z}(\boldsymbol{X})\big)=\frac{1}{(\sqrt{2\pi})^n}e^{-\frac{1}{2}\boldsymbol{Z}^T\boldsymbol{Z}}$，其中指数部分

$$\boldsymbol{Z}^T\boldsymbol{Z}=\left(\boldsymbol{A}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right)^T\boldsymbol{A}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})=(\boldsymbol{X}-\boldsymbol{\mu})^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}(\boldsymbol{X}-\boldsymbol{\mu})$$

而多元函数换元雅克比行列式$\mathcal{J}=\left|\frac{dz}{dx}\right|=|\boldsymbol{A}^{-1}|=|\boldsymbol{A}|^{-1}=|\boldsymbol{A}^T|^{-1}=|\boldsymbol{A}\boldsymbol{A}^T|^{-\frac{1}{2}}$

概率密度积分变为

$$1=\int_{-\infty}^{+\infty}\frac{1}{(\sqrt{2\pi})^n}e^{-\frac{1}{2}\boldsymbol{Z}^T\boldsymbol{Z}}d\boldsymbol{z}=\int_{-\infty}^{+\infty}\frac{1}{(\sqrt{2\pi})^n|\boldsymbol{A}\boldsymbol{A}^T|^{\frac{1}{2}}}e^{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})^T\left(\boldsymbol{A}\boldsymbol{A}^T\right)^{-1}(\boldsymbol{X}-\boldsymbol{\mu})}\,d\boldsymbol{x}$$

协方差矩阵:

$\boldsymbol{\Sigma}=\mathbb{E}[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T]$
$=\mathbb{E}[(\boldsymbol{A}\boldsymbol{Z})(\boldsymbol{A}\boldsymbol{Z})^T]=\boldsymbol{A}cov(\boldsymbol{Z},\boldsymbol{Z})\boldsymbol{A}^T=$
$\boldsymbol{A}\boldsymbol{A}^T$ (半正定矩阵)

https://zhuanlan.zhihu.com/p/58987388

01 概率分布

- **不同模型的函数传播**

■ 仿射矩阵变换(线性模型)

➢ 已知多维随机变量$\boldsymbol{X}$的一阶矩$\boldsymbol{\mu}$和二阶矩$\boldsymbol{\Sigma}$

➢ 仿射矩阵变换$\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}$

➢ 随机变量$\boldsymbol{Y}$的一阶矩$\boldsymbol{\mu}_{\boldsymbol{y}}=\boldsymbol{A}\boldsymbol{\mu}+\boldsymbol{b}$，二阶矩$\boldsymbol{\Sigma}_{\boldsymbol{y}}=\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T$

■ 非线性变换(非线性模型)

➢ 已知多维随机变量$\boldsymbol{X}$的概率密度分布为$f(\boldsymbol{x})$

➢ 仿射矩阵变换$\boldsymbol{y}=\boldsymbol{g}(\boldsymbol{x})$

➢ 随机变量$\boldsymbol{Y}$的概率密度分布为$f(\boldsymbol{y})=f(\boldsymbol{x})\left|\frac{d\boldsymbol{g}}{d\boldsymbol{x}}\right|^{-1}$

■ 卷积变换(卷积模型)

➢ 两个相互独立的随机变量$\boldsymbol{X}_1$和$\boldsymbol{X}_2$的一阶矩分别为$\boldsymbol{\mu}_1$和$\boldsymbol{\mu}_2$，二阶矩分别为$\boldsymbol{\Sigma}_1$和$\boldsymbol{\Sigma}_2$

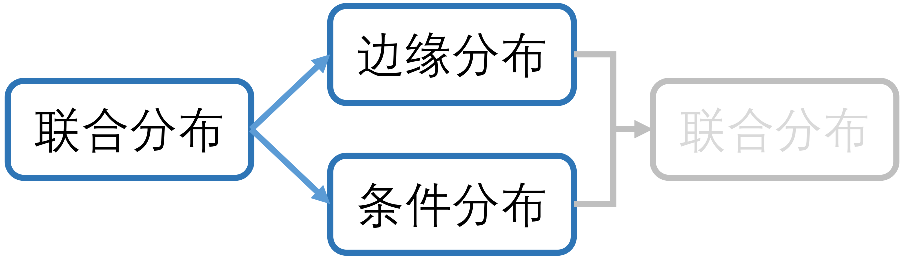
➢ 叠加变换$\boldsymbol{y}=\boldsymbol{x}_1+\boldsymbol{x}_2$

➢ 随机变量$\boldsymbol{Y}$的一阶矩为$\boldsymbol{\mu}_{\boldsymbol{y}}=\boldsymbol{\mu}_1+\boldsymbol{\mu}_2$，二阶矩为$\boldsymbol{\Sigma}_{\boldsymbol{y}}=\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2$

➢ 高斯分布的相互独立与线性无关是等价的

注：
1. 高斯分布的线性传播仍为高斯分布
2. 高斯分布的非线性传播不一定为高斯分布
3. 相互高斯分布的叠加分布仍为高斯分布

● 联合分布−>边缘分布与条件分布



■ 基本关系

- 联合分布可以分别推导出边缘分布和条件分布
- 边缘分布和条件分布共同才能推出联合分布

■ 高斯分布的边缘分布与条件分布

- 已知n维随机变量$\boldsymbol{X}=[\boldsymbol{X}_1,\boldsymbol{X}_2]^T\sim\mathcal{N}(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$
- $\boldsymbol{X}_1$为p维随机向量，$\boldsymbol{X}_2$为q维随机向量；$p+q=n$
- 矩阵可以分解为 $\boldsymbol{\mu}=\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix}$，$\boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}$($\boldsymbol{\Sigma}$为半正定矩阵，即$\boldsymbol{\Sigma}_{21}=\boldsymbol{\Sigma}_{12}^T$)
- 可以得到如下两个结论：(高斯模型proof2)
- (1) 边缘分布$\boldsymbol{X}_1$和$\boldsymbol{X}_2$均为高斯分布，且$\boldsymbol{X}_1\sim\mathcal{N}(\boldsymbol{\mu}_1,\ \boldsymbol{\Sigma}_{11})$ 公式1，$\boldsymbol{X}_2\sim\mathcal{N}(\boldsymbol{\mu}_2,\ \boldsymbol{\Sigma}_{22})$ 公式2
- (2) 条件分布$\boldsymbol{X}_1|\boldsymbol{X}_2$和$\boldsymbol{X}_2|\boldsymbol{X}_1$均为高斯分布，且

  $\boldsymbol{X}_1|\boldsymbol{X}_2\sim\mathcal{N}(\boldsymbol{\mu}_1+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}_2-\boldsymbol{\mu}_2),\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$ 公式3

  $\boldsymbol{X}_2|\boldsymbol{X}_1\sim\mathcal{N}(\boldsymbol{\mu}_2+\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X}_1-\boldsymbol{\mu}_1),\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})$ 公式4
- (1)(2)可以融合成综合的公式：

$$\boldsymbol{X}_j|\boldsymbol{X}_i\sim\mathcal{N}\left(\boldsymbol{\mu}_j+\delta_{ij}\boldsymbol{\Sigma}_{ji}\boldsymbol{\Sigma}_{ii}^{-1}(\boldsymbol{X}_i-\boldsymbol{\mu}_i),\ \boldsymbol{\Sigma}_{jj}-\delta_{ij}\boldsymbol{\Sigma}_{ji}\boldsymbol{\Sigma}_{ii}^{-1}\boldsymbol{\Sigma}_{ij}\right)$$

其中$\delta_{ij}=1*(i\neq j)+0*(i=j)$

当$i=j=1$时，公式1
当$i=j=2$时，公式2
当$i=1,j=2$时，公式3
当$i=2,j=1$时，公式4

- Schur补定理： ➢ ($\boldsymbol{\Sigma}$为半正定矩阵，即$\boldsymbol{\Sigma}_{21}=\boldsymbol{\Sigma}_{12}^T$)
- $\boldsymbol{\Sigma}^{11}=(\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^T)^{-1}=\boldsymbol{\Sigma}_{11}^{-1}+\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})\boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}$
- $\boldsymbol{\Sigma}^{22}=(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}=\boldsymbol{\Sigma}_{22}^{-1}+\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^T(\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}^T)\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$
- $\boldsymbol{\Sigma}^{12}=-\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}=(\boldsymbol{\Sigma}^{12})^T$

- **Proof2(联合到条件)**

联合概率密度函数：$f(\boldsymbol{X}_1,\boldsymbol{X}_2)=\frac{1}{(\sqrt{2\pi})^n|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right)=\frac{1}{(\sqrt{2\pi})^n|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\boldsymbol{Q}(\boldsymbol{X}_1,\boldsymbol{X}_2)\right)$

考虑$\boldsymbol{Q}(\boldsymbol{X}_1,\boldsymbol{X}_2)=(\boldsymbol{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})=[(\boldsymbol{X}_1-\boldsymbol{\mu}_1)^T,(\boldsymbol{X}_2-\boldsymbol{\mu}_2)^T]\begin{bmatrix}\boldsymbol{\Sigma}^{11} & \boldsymbol{\Sigma}^{12}\\ \boldsymbol{\Sigma}^{21} & \boldsymbol{\Sigma}^{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_1-\boldsymbol{\mu}_1\\ \boldsymbol{X}_2-\boldsymbol{\mu}_2\end{bmatrix}$

$=(\boldsymbol{X}_1-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{11}(\boldsymbol{X}_1-\boldsymbol{\mu}_1)+2(\boldsymbol{X}_1-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{12}(\boldsymbol{X}_2-\boldsymbol{\mu}_2)+(\boldsymbol{X}_2-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{22}(\boldsymbol{X}_2-\boldsymbol{\mu}_2)$

代入Schur补定理：

$\boldsymbol{Q}(\boldsymbol{X}_1,\boldsymbol{X}_2)=(\boldsymbol{X}_1-\boldsymbol{\mu}_1)^T[\boldsymbol{\Sigma}_{11}^{-1}+\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})\boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}](\boldsymbol{X}_1-\boldsymbol{\mu}_1)$

$-2(\boldsymbol{X}_1-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}(\boldsymbol{X}_2-\boldsymbol{\mu}_2)$

$+(\boldsymbol{X}_2-\boldsymbol{\mu}_2)^T[(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}](\boldsymbol{X}_2-\boldsymbol{\mu}_2)$

$=(\boldsymbol{X}_1-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X}_1-\boldsymbol{\mu}_1)$

$+(\boldsymbol{X}_1-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})\boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X}_1-\boldsymbol{\mu}_1)$

$-2(\boldsymbol{X}_1-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}(\boldsymbol{X}_2-\boldsymbol{\mu}_2)$

$+(\boldsymbol{X}_2-\boldsymbol{\mu}_2)^T[(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}](\boldsymbol{X}_2-\boldsymbol{\mu}_2)$

二次项定理：$A=A^T$

$u^TAu-2u^TAv+v^TAv=u^TAu-u^TAv-u^TAv+v^TAv$

$=u^TA\,(u-v)-(u-v)^TAv$

$=u^TA\,(u-v)-v^TA^T(u-v)$

$=(u-v)^TA(u-v)$

$=(\boldsymbol{X}_1-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X}_1-\boldsymbol{\mu}_1)+[(\boldsymbol{X}_2-\boldsymbol{\mu}_2)-\boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X}_1-\boldsymbol{\mu}_1)]^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}[(\boldsymbol{X}_2-\boldsymbol{\mu}_2)-\boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X}_1-\boldsymbol{\mu}_1)]$

- Schur补定理：
- $|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}_{11}|\left|\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right|$

- **Proof2(联合到条件)**

定义：$\boldsymbol{b} := \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X}_1 - \boldsymbol{\mu}_1)$, $\boldsymbol{A} := \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$，则：

- $\boldsymbol{Q}_1(\boldsymbol{X}_1) := (\boldsymbol{X}_1 - \boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X}_1 - \boldsymbol{\mu}_1)$
- $\boldsymbol{Q}_2(\boldsymbol{X}_1, \boldsymbol{X}_2) := (\boldsymbol{X}_2 - \boldsymbol{b})^T\boldsymbol{A}^{-1}(\boldsymbol{X}_2 - \boldsymbol{b})$
- $\boldsymbol{Q}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \boldsymbol{Q}_1(\boldsymbol{X}_1) + \boldsymbol{Q}_2(\boldsymbol{X}_1, \boldsymbol{X}_2)$

考虑联合概率密度

$$f(\boldsymbol{X}_1, \boldsymbol{X}_2) = \frac{1}{\left(\sqrt{2\pi}\right)^n|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\boldsymbol{Q}(\boldsymbol{X}_1, \boldsymbol{X}_2)\right)$$

$$= \frac{1}{\left(\sqrt{2\pi}\right)^{p/2}|\boldsymbol{\Sigma}_{11}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\boldsymbol{X}_1 - \boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X}_1 - \boldsymbol{\mu}_1)\right)\frac{1}{\left(\sqrt{2\pi}\right)^{q/2}|\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\boldsymbol{X}_2 - \boldsymbol{b})^T\boldsymbol{A}^{-1}(\boldsymbol{X}_2 - \boldsymbol{b})\right)$$

$$= \mathcal{N}(\boldsymbol{X}_1, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})\mathcal{N}(\boldsymbol{X}_2, \boldsymbol{b}, \boldsymbol{A})$$

边缘概率密度$f_1(\boldsymbol{X}_1) = \int_{-\infty}^{+\infty} f(\boldsymbol{X}_1, \boldsymbol{X}_2)d\boldsymbol{X}_2 = \frac{1}{\left(\sqrt{2\pi}\right)^{p/2}|\boldsymbol{\Sigma}_{11}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\boldsymbol{X}_1 - \boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X}_1 - \boldsymbol{\mu}_1)\right)$

条件概率密度$f_{2|1}(\boldsymbol{X}_2|\boldsymbol{X}_1) = \frac{f(\boldsymbol{X}_1, \boldsymbol{X}_2)}{f_1(\boldsymbol{X}_1)} = \frac{1}{\left(\sqrt{2\pi}\right)^{q/2}|\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\boldsymbol{X}_2 - \boldsymbol{b})^T\boldsymbol{A}^{-1}(\boldsymbol{X}_2 - \boldsymbol{b})\right)$

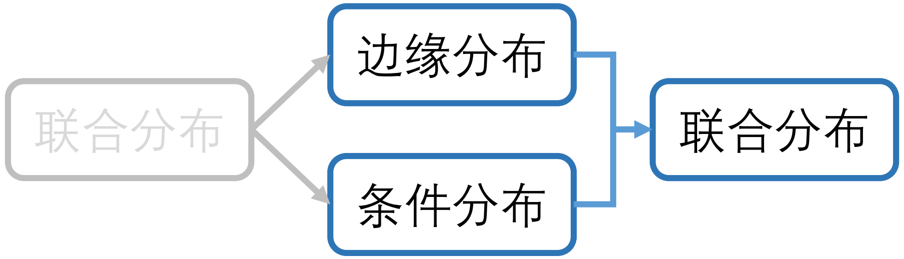
从而可知：$\boldsymbol{X}_2|\boldsymbol{X}_1 \sim \mathcal{N}(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X}_1 - \boldsymbol{\mu}_1),\ \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})$

● 边缘分布与条件分布−>联合分布

■ 基本关系

➢ 联合分布可以分别推导出边缘分布和条件分布

➢ 边缘分布和条件分布共同才能推出联合分布



■ 高斯分布的边缘分布与条件分布

➢ 已知$\boldsymbol{X}_1$为p维高斯分布，$\boldsymbol{X}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$；$\boldsymbol{X}_2|\boldsymbol{X}_1$为$q$维高斯分布，$\boldsymbol{X}_2|\boldsymbol{X}_1 \sim \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$;

➢ n维随机变量$\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2]^T$的一阶矩为$\boldsymbol{\mu}$，二阶矩为$\boldsymbol{\Sigma}$

➢ 可以得到如下结论：(乘积模型)

➢ 联合分布$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,

➢ 存在两种情况

a) $\boldsymbol{X}_1$与$\boldsymbol{X}_2$相互独立，则$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$， $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{22} \end{bmatrix} (p + q = n)$

b) $\boldsymbol{X}_1$与$\boldsymbol{X}_2$相关，由proof2可知必定为线性关系，不妨设线性关系为$\boldsymbol{X}_2 = \boldsymbol{A}\boldsymbol{X}_1 + \boldsymbol{b}$，则$\boldsymbol{X}_2|\boldsymbol{X}_1 \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{X}_1 + \boldsymbol{b}, \boldsymbol{\Sigma}_{22})$,

$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{A}\boldsymbol{\mu}_1 + \boldsymbol{b} \end{bmatrix}$， $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{11}\boldsymbol{A}^T \\ \boldsymbol{A}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{22} + \boldsymbol{A}\boldsymbol{\Sigma}_{11}\boldsymbol{A}^T \end{bmatrix}$

两个式子均可以统一为b)中的式子

http://fourier.eng.hmc.edu/e161/lectures/gaussianprocess/node7.html

01 概率分布

- **Proof3(条件到联合)**

联合概率密度函数:

$$f(\boldsymbol{X}_1,\boldsymbol{X}_2)=f_{2|1}(\boldsymbol{X}_2|\boldsymbol{X}_1)f_1(\boldsymbol{X}_1)=\frac{1}{\left(\sqrt{2\pi}\right)^{n/2}|\boldsymbol{\Sigma}_{11}|^{\frac{1}{2}}|\boldsymbol{\Sigma}_{22}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\boldsymbol{X}_1-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X}_1-\boldsymbol{\mu}_1)-\frac{1}{2}(\boldsymbol{X}_2-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}_2-\boldsymbol{\mu}_2)\right)$$

考虑指数部分: $\boldsymbol{Q}(\boldsymbol{X}_1,\boldsymbol{X}_2)=(\boldsymbol{X}_1-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X}_1-\boldsymbol{\mu}_1)+\left(\boldsymbol{X}_2-(\boldsymbol{A}\boldsymbol{X}_1+\boldsymbol{b})\right)^T\boldsymbol{\Sigma}_{22}^{-1}\left(\boldsymbol{X}_2-(\boldsymbol{A}\boldsymbol{X}_1+\boldsymbol{b})\right)$

展开可得: $\boldsymbol{Q}(\boldsymbol{X}_1,\boldsymbol{X}_2)=\boldsymbol{X}_1^T(\boldsymbol{\Sigma}_{11}^{-1}+\boldsymbol{A}^T\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{A})\boldsymbol{X}_1-2\boldsymbol{X}_1^T\left(\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1+\boldsymbol{A}^T\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}_2-\boldsymbol{b})\right)+(\boldsymbol{X}_2-\boldsymbol{b})^T\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}_2-\boldsymbol{b})+\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1$

配方可得: $\boldsymbol{Q}(\boldsymbol{X}_1,\boldsymbol{X}_2)=\left[(\boldsymbol{X}_1-\boldsymbol{\mu}_1)^T,\left(\boldsymbol{X}_2-(\boldsymbol{A}\boldsymbol{\mu}_1+\boldsymbol{b})\right)^T\right]\begin{bmatrix}\boldsymbol{\Sigma}_{11}^{-1}+\boldsymbol{A}^T\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{A} & -\boldsymbol{A}^T\boldsymbol{\Sigma}_{22}^{-1}\\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{A} & \boldsymbol{\Sigma}_{22}^{-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_1-\boldsymbol{\mu}_1\\ \boldsymbol{X}_2-(\boldsymbol{A}\boldsymbol{\mu}_1+\boldsymbol{b})\end{bmatrix}+\boldsymbol{C}$

从而可得: $\boldsymbol{\mu}=\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{A}\boldsymbol{\mu}_1+\boldsymbol{b}\end{bmatrix}$, $\boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_{11}^{-1}+\boldsymbol{A}^T\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{A} & -\boldsymbol{A}^T\boldsymbol{\Sigma}_{22}^{-1}\\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{A} & \boldsymbol{\Sigma}_{22}^{-1}\end{bmatrix}^{-1}=\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{11}\boldsymbol{A}^T\\ \boldsymbol{A}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{22}+\boldsymbol{A}\boldsymbol{\Sigma}_{11}\boldsymbol{A}^T\end{bmatrix}$

结论:

(1) 联合分布$\boldsymbol{X}\sim\mathcal{N}(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$

(2) 边缘分布$\boldsymbol{X}_2\sim\mathcal{N}(\boldsymbol{A}\boldsymbol{\mu}_1+\boldsymbol{b},\boldsymbol{\Sigma}_{22}+\boldsymbol{A}\boldsymbol{\Sigma}_{11}\boldsymbol{A}^T)$

● 高斯分布总结

■ 卷积分布

- 相互独立的随机变量$\boldsymbol{X}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$，$\boldsymbol{X}_2 \sim \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$
- 叠加变换$\boldsymbol{y} = \boldsymbol{x}_1 + \boldsymbol{x}_2$
- 随机变量$\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)$

■ 联合分布−>边缘分布与条件分布

- 已知n维随机变量$\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2]^T \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，可分解为$\boldsymbol{X}_1$为$p$维随机向量，$\boldsymbol{X}_2$为$q$维随机向量；$p + q = n$
- 其中 $\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$， $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$($\boldsymbol{\Sigma}$为半正定矩阵，即$\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}^T$)
- 则$\boldsymbol{X}_j | \boldsymbol{X}_i \sim \mathcal{N}\left(\boldsymbol{\mu}_j + \delta_{ij} \boldsymbol{\Sigma}_{ji} \boldsymbol{\Sigma}_{ii}^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu}_i), \boldsymbol{\Sigma}_{jj} - \delta_{ij} \boldsymbol{\Sigma}_{ji} \boldsymbol{\Sigma}_{ii}^{-1} \boldsymbol{\Sigma}_{ij}\right)$，其中$\delta_{ij} = 1 * (i \neq j) + 0 * (i = j)$

■ 边缘分布与条件分布−> 联合分布

- $\boldsymbol{X}_1$为p维高斯分布，$\boldsymbol{X}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$；$\boldsymbol{X}_2 | \boldsymbol{X}_1$为$q$维高斯分布，$\boldsymbol{X}_2 | \boldsymbol{X}_1 \sim \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$
- 若满足$\boldsymbol{X}_2 | \boldsymbol{X}_1 \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{X}_1 + \boldsymbol{b}, \boldsymbol{\Sigma}_{22})$，线性无关则$\boldsymbol{A} = 0$
- 联合分布$\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2]^T$服从高斯分布，$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{A}\boldsymbol{\mu}_1 + \boldsymbol{b} \end{bmatrix}$， $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{11} \boldsymbol{A}^T \\ \boldsymbol{A}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{22} + \boldsymbol{A}\boldsymbol{\Sigma}_{11}\boldsymbol{A}^T \end{bmatrix}$
- 边缘分布$\boldsymbol{X}_2 \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{\mu}_1 + \boldsymbol{b}, \boldsymbol{\Sigma}_{22} + \boldsymbol{A}\boldsymbol{\Sigma}_{11}\boldsymbol{A}^T)$