关于多维正态分布

定义 设 $\mu \in \mathrm{R}^n$，Σ 是 n 阶实对称正定方阵，称 n 维随机向量 X 服从正态分布 $N(\mu,\Sigma)$，如果 X 有以下形式的联合概率密度函数

$$f_X(x)=\frac{1}{\sqrt{(2\pi)^n\det(\Sigma)}}\exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right)$$

教材相关内容：第 180 页例 3.4.12。$n=2$ 的情形，第 141 页二元正态分布。

性质 1：（正态分布在可逆仿射变换下仍是正态分布） 设 n 维随机向量 $X\sim N(\mu,\Sigma)$，A 是 n 阶实数可逆方阵，$b\in\mathrm{R}^n$。则 $Y=AX+b\sim N(A\mu+b, A\Sigma A')$。

证明：注意到

$$x=A^{-1}(Y-b)\text{，}\quad x-\mu=A^{-1}(Y-A\mu-b)\text{，}$$

$$\det(A\Sigma A')=\det(A)\cdot\det(\Sigma)\cdot\det(A')=\det(\Sigma)\cdot\det(A)^2$$

所以，

$$\begin{aligned}
f_Y(y)&=f_{AX+b}(y)\\
&=f_X(A^{-1}(y-b))\times\frac{1}{|\det A|}\\
&=\frac{1}{\sqrt{(2\pi)^n\det\Sigma}}\exp\left(-\frac{1}{2}(A^{-1}(y-b)-\mu)'\Sigma^{-1}(A^{-1}(y-b)-\mu)\right)\times\frac{1}{|\det A|}\\
&=\frac{1}{\sqrt{(2\pi)^n\det(A\Sigma A')}}\exp\left(-\frac{1}{2}(y-b-A\mu)'(A\Sigma A')^{-1}(y-b-A\mu)\right)
\end{aligned}$$

故 $Y=AX+b\sim N(A\mu+b, A\Sigma A')$。

教材相关内容：第 162 页例 3.3.9。

性质 2：（具有独立分量的正态分布随机向量，边缘分布）
设

$$X=\begin{pmatrix}X_1\\X_2\end{pmatrix}\sim N\left(\begin{pmatrix}\mu_1\\\mu_2\end{pmatrix},\begin{pmatrix}\Sigma_{11}&0\\0&\Sigma_{22}\end{pmatrix}\right)$$

其中 X 是 n 维随机向量，X_1 是 n_1 维随机向量，X_2 是 n_2 维随机向量；$\mu_i\in\mathrm{R}^{n_i}$，Σ_{ii} 是 n_i 阶实数矩阵，$i=1,2$。则 Σ_{ii} 是对称正定矩阵，X_1 与 X_2 独立，并且 $X_i\sim N(\mu_i,\Sigma_{ii})$，$i=1,2$。

证明：1、易见 Σ_{ii} 是对称矩阵，$i=1,2$。

$$x_1'\Sigma_{11}x_1=(x_1'\quad 0)\begin{pmatrix}\Sigma_{11}&0\\0&\Sigma_{22}\end{pmatrix}\begin{pmatrix}x_1\\0\end{pmatrix}\geq 0$$

而且 $x_1'\Sigma_{11}x_1=0$ 当且仅当 $x_1=0$。因此 Σ_{11} 是对称正定矩阵。类似可证 Σ_{22} 是对称正定矩阵。

2、对

$$\Sigma=\begin{pmatrix}\Sigma_{11}&0\\0&\Sigma_{22}\end{pmatrix},$$

自然有

$$\Sigma^{-1}=\begin{pmatrix}\Sigma_{11}^{-1}&0\\0&\Sigma_{22}^{-1}\end{pmatrix},\qquad \det\Sigma=\det\Sigma_{11}\cdot\det\Sigma_{22}$$

从而

$$\begin{aligned}f_{X_1,X_2}(x_1,x_2)&=\frac{1}{\sqrt{(2\pi)^n\det\Sigma}}\exp\left(-\frac{1}{2}(x_1',x_2')\Sigma^{-1}\begin{pmatrix}x_1\\x_2\end{pmatrix}\right)\\&=\prod_{i=1}^{2}\frac{1}{\sqrt{(2\pi)^{n_i}\det\Sigma_{ii}}}\exp\left(-\frac{1}{2}x_i'\Sigma_{ii}^{-1}x_i\right)\end{aligned}$$

易见

$$f_{X_i}(x_i)=\frac{1}{\sqrt{(2\pi)^{n_i}\det\Sigma_{ii}}}\exp\left(-\frac{1}{2}x_i'\Sigma_{ii}^{-1}x_i\right),\qquad i=1,2.$$

从而

$$f_{X_1,X_2}(x_1,x_2)=f_{X_1}(x_1)\cdot f_{X_2}(x_2)$$

故 X_1,X_2 独立，且 $X_i\sim N(\mu_i,\Sigma_{ii})$。

性质 3：（正态分布随机向量的分量的独立化，正态分布的边缘分布仍是正态分布）

设

$$X=\begin{pmatrix}X_1\\X_2\end{pmatrix}\sim N\left(\begin{pmatrix}\mu_1\\\mu_2\end{pmatrix},\begin{pmatrix}\Sigma_{11}&\Sigma_{12}\\\Sigma_{21}&\Sigma_{22}\end{pmatrix}\right)$$

其中 X 是 n 维随机向量，X_1 是 n_1 维随机向量，X_2 是 n_2 维随机向量；$\mu_i\in\mathrm{R}^{n_i}$，Σ_{ij} 是 $n_i\times n_j$ 阶实数矩阵，$i=1,2$。记

$$Y=\begin{pmatrix}Y_1\\Y_2\end{pmatrix}=\begin{pmatrix}I_1&0\\-\Sigma_{21}\Sigma_{11}^{-1}&I_2\end{pmatrix}\begin{pmatrix}X_1\\X_2\end{pmatrix}$$

其中 I_i 是 n_i 阶单位矩阵。则

1. $Y\sim N\left(\begin{pmatrix}\mu_1\\\mu_2-\Sigma_{21}\Sigma_{11}^{-1}\mu_1\end{pmatrix},\begin{pmatrix}\Sigma_{11}&0\\0&\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\end{pmatrix}\right)$，从而 Y_1 和 Y_2 独立。

2. $X_i\sim N(\mu_i,\Sigma_{ii})$，$i=1,2$。

证明：根据性质 1，

$$\begin{aligned}Y&=\begin{pmatrix}Y_1\\Y_2\end{pmatrix}=\begin{pmatrix}I_1&0\\-\Sigma_{21}\Sigma_{11}^{-1}&I_2\end{pmatrix}\begin{pmatrix}X_1\\X_2\end{pmatrix}\\&\sim N\left(\begin{pmatrix}I_1&0\\-\Sigma_{21}\Sigma_{11}^{-1}&I_2\end{pmatrix}\begin{pmatrix}\mu_1\\\mu_2\end{pmatrix},\begin{pmatrix}I_1&0\\-\Sigma_{21}\Sigma_{11}^{-1}&I_2\end{pmatrix}\begin{pmatrix}\Sigma_{11}&\Sigma_{12}\\\Sigma_{21}&\Sigma_{22}\end{pmatrix}\begin{pmatrix}I_1&0\\-\Sigma_{21}\Sigma_{11}^{-1}&I_2\end{pmatrix}\right)\\&=N\left(\begin{pmatrix}\mu_1\\\mu_2-\Sigma_{21}\Sigma_{11}^{-1}\mu_1\end{pmatrix},\begin{pmatrix}\Sigma_{11}&0\\0&\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\end{pmatrix}\right)\end{aligned}$$

由性质 2 知，Y_1 和 Y_2 独立，并且 $X_1=Y_1\sim N(\mu_1,\Sigma_{11})$。用类似的办法可以证明 $X_2\sim N(\mu_2,\Sigma_{22})$。

注记：这里使用的变量变换是从不独立（X_1,X_2 可能不独立）到独立（构造出来的 Y_1,Y_2 是独立的），而对正态分布随机向量的分量，独立与不相关是等价的（性质 3），而不相关相当于几何上的垂直（关于 H^2 空间上的内积），因此这本质上就是内积空间中向量组的 **Gram-Schmidt** 正交化过程。这方法在教材第 141 页例 3.1.7、第 149 页例 3.2.5、第 174 页例 3.4.9、第 189 页例 3.5.4 中都有体现。另外，这里得到的结论对应教材第 149 页例 3.2.5（二元正态的边缘分布）。

性质 1'：（正态分布在非退化仿射变换下的不变性，性质 1 的一般形式）

设 n 维随机向量 $X \sim N(\mu,\Sigma)$，A 是 $m\times n$ 阶实数方阵，$\mathrm{rank}A=m$（即 A 的行向量是线性无关的），$b\in \mathrm{R}^m$。则 $Y=AX+b\sim N(A\mu+b, A\Sigma A')$。

证明：因为 A 是满行秩矩阵，所以 $m\le n$。如果 $m=n$，则 A 是可逆矩阵，这时结论如 b 中形式。

如果 $m<n$，则 A 的 m 个 n 维行向量线性无关，我们可以将它们扩充为 n 维空间的一组基，也就是说存在 $(n-m)\times n$ 矩阵 B 使得

$$\begin{pmatrix} A \\ B \end{pmatrix}_{n\times n}$$

是可逆矩阵，由性质 1，

$$\begin{pmatrix} Y \\ Z \end{pmatrix}=\begin{pmatrix} A \\ B \end{pmatrix}X+\begin{pmatrix} b \\ 0 \end{pmatrix}\sim N\left(\begin{pmatrix} A\mu+b \\ B\mu \end{pmatrix},\begin{pmatrix} A \\ B \end{pmatrix}\Sigma(A',B')\right)=N\left(\begin{pmatrix} A\mu+b \\ B\mu \end{pmatrix},\begin{pmatrix} A\Sigma A' & A\Sigma B' \\ B\Sigma A' & B\Sigma B' \end{pmatrix}\right)$$

由性质 3 知，它的一个边缘分布为 $Y=AX+b\sim N(A\mu+b, A\Sigma A')$。

一个常用的结论（性质 1'的特例）

设 $X_1,X_2,\dots,X_n$ 独立，$X_i\sim N(\mu_i,\sigma_i^2)$，$i=1,2,\dots,n$，$a_1,a_2,\dots,a_n,b\in\mathrm{R}$，其中 $a_1,a_2,\dots,a_n$ 不全为零。则

$$a_1X_1+\cdots a_nX_n+b\sim N(a_1\mu_1+\cdots a_n\mu_n+b, a_1^2\sigma_1^2+\cdots+a_n^2\sigma_n^2)。$$

证明：由独立性知，

$$f_{X_1,\dots,X_n}(x_1,\cdots,x_n)=f_{X_1}(x_1)\cdots f_{X_n}(x_n)=\frac{1}{\sqrt{(2\pi)^n\sigma_1^2\cdots\sigma_n^2}}\exp\left(-\frac{1}{2}\sum_{i=1}^n\frac{(x_i-\mu_i)^2}{\sigma_i^2}\right)$$

即

$$(X_1,\dots,X_n)'\sim N\left(\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix},\begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{pmatrix}\right),$$

于是在性质 5 中取 $A=(a_1,\dots,a_n)$。因 $a_1,a_2,\dots,a_n$ 不全为零，故 A 满行秩。于是应用性质 1'就得到这个结论。

教材相关内容： 第 159 页例 3.3.6。

性质 4:（正态分布参数的概率含义） 设n维随机向量$X \sim N(\mu,\Sigma)$。则$E(X)=\mu$（即$E(X_i)=\mu_i$，$i=1,\dots,n$），Σ是X的协方差矩阵（即$\Sigma_{i,j}=\mathrm{Cov}(X_i,X_j)$，$i,j=1,\dots,n$）。

证明：因为Σ是n阶实对称正定方阵，所以存在n阶正交矩阵C使得

$$C'\Sigma C=\begin{pmatrix}\lambda_1^2 & & \\ & \ddots & \\ & & \lambda_n^2\end{pmatrix},\quad \lambda_1>0,\dots,\lambda_n>0.$$

记

$$\Lambda=\begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{pmatrix},\quad A=C\Lambda,$$

则A是n阶可逆矩阵，$C'\Sigma C=\Lambda^2$，于是$\Sigma=C\Lambda^2C'=AA'$。

令$Y=A^{-1}(X-\mu)$，则由性质 1 知道

$$Y=A^{-1}X-A^{-1}\mu\sim N(A^{-1}\mu-A^{-1}\mu,A^{-1}\Sigma A^{-1'})=N(0,\mathrm{I}_n)$$

其中I_n是n阶单位矩阵。于是Y的联合概率密度函数为

$$f_Y(y)=\frac{1}{\sqrt{(2\pi)^n}}\exp\left(-\frac{1}{2}y'y\right)=\prod_{k=1}^{n}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y_k^2\right),$$

因此$Y_1,\dots,Y_n\overset{i.i.d.}{\sim}N(0,1)$，于是

$$E(Y_k)=0,\quad \mathrm{Cov}(Y_i,Y_j)=\begin{cases}1, & i=j;\\ 0, & i\neq j.\end{cases}，即 E(Y)=0,\quad \Sigma_Y=\mathrm{I}_n,$$

因此，由数学期望和协方差矩阵的性质，得到

$$E(X)=E(AY+\mu)=AE(Y)+\mu=\mu,\quad \Sigma_X=A\Sigma_Y A'=AA'=\Sigma。$$

上述证明给出的构造过程叫做正态分布的标准化过程（我们称$N(0,\mathrm{I}_n)$为n维标准正态分布）。

教材相关内容： 有了协方差矩阵就可以很容易地得到相关系数，对$n=2$的情形，

$$\Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix},$$

所以

$$\mathrm{Var}(X_i)=\sigma_i^2,\quad \mathrm{Cov}(X_1,X_2)=\rho\sigma_1\sigma_2,$$

从而

$$\rho=\frac{\rho\sigma_1\sigma_2}{\sigma_1\sigma_2}=\frac{\mathrm{Cov}(X_1,X_2)}{\sqrt{\mathrm{Var}X_1\cdot\mathrm{Var}X_2}}=\rho_{X_1,X_2}$$

是 X_1,X_2 的相关系数，这就是教材第 174 页例 3.4.9 的结论。

教材第 180 页例 3.4.12 中定义多维正态分布时，数学期望向量和协方差矩阵的说法本不应该写在定义的叙述中，因为它们是概率密度函数的自然推论。

性质 5：（对正态分布随机向量的分量，独立性与不相关性等价）

设

$$X=\begin{pmatrix}X_1\\X_2\end{pmatrix}\sim N\left(\begin{pmatrix}\mu_1\\\mu_2\end{pmatrix},\begin{pmatrix}\Sigma_{11}&\Sigma_{12}\\\Sigma_{21}&\Sigma_{22}\end{pmatrix}\right)$$

其中 X 是 n 维随机向量，X_1 是 n_1 维随机向量，X_2 是 n_2 维随机向量；$\mu_i\in\mathrm{R}^{n_i}$，Σ_{ij} 是 $n_i\times n_j$ 阶实数矩阵，$i=1,2$。则 X_1 与 X_2 独立当且仅当 $\Sigma_{12}=\Sigma_{21}'=0$；

证明：充分性就是性质 2。下证必要性。因 X_1 与 X_2 独立，所以 X_1 的任何分量 U 与 X_2 的任何分量 V 独立，于是 $\mathrm{Cov}(U,V)=0$，而根据性质 4，Σ 是 X 的协方差矩阵，$\Sigma_{12}=\Sigma_{21}'$ 的元素是 X_1 的分量与 X_2 的分量的协方差，因此 $\Sigma_{12}=\Sigma_{21}'=0$。

教材相关内容： 第 178 页性质 3.4.13。

性质 6：（正态分布的条件分布仍是正态分布）
设

$$X=\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}\sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)。$$

则

1. 在已知 $X_1=x_1$ 发生的条件下，X_2 的条件概率分布是正态分布

$$N(\mu_2+\Sigma_{21}\Sigma_{11}^{-1}(x_1-\mu_1),\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$$

2. X_2 关于 X_1 的线性回归与非线性回归相同：$E(X_2\,|\,X_1)=\mu_2+\Sigma_{21}\Sigma_{11}^{-1}(X_1-\mu_1)$。

证明：由性质 3，我们知道 $Y_2=X_2-\Sigma_{21}\Sigma_{11}^{-1}X_1$ 与 $Y_1=X_1$ 独立，并且

$$Y_2\sim N(\mu_2-\Sigma_{21}\Sigma_{11}^{-1}\mu_1,\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$$

于是在已知 $X_1=x_1$ 的条件下，$X_2=Y_2+\Sigma_{21}\Sigma_{11}^{-1}x_1$，$X_2$ 的条件概率分布就是 $Y_2+\Sigma_{21}\Sigma_{11}^{-1}x_1$ 的概率分布，根据性质 1，$Y_2+\Sigma_{21}\Sigma_{11}^{-1}x_1$ 的分布是正态分布

$$N(\mu_2+\Sigma_{21}\Sigma_{11}^{-1}(x_1-\mu_1),\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}),$$

这就是在已知 $X_1=x_1$ 的条件下 X_2 的条件概率分布。再有性质 4 知，这个条件分布的数学期望为

$$E(X_2\,|\,X_1=x_1)=\mu_2+\Sigma_{21}\Sigma_{11}^{-1}(x_1-\mu_1)。$$

由于它关于 x_1 是一次的，所以这即是线性回归又是非线性回归。

教材相关内容： 第 189 页例 3.5.4。第 193 页第 13 行的公式。