多元正态分布的定义

目录

定义 1: 通过概率密度函数 (PDF) (当协方差矩阵 Σ 正定时)

这是最直观和常见的一种定义，但它有一个前提条件：

> 协方差矩阵 Σ 必须是正定的 (positive definite)，即 $|\Sigma| > 0$。

> 一个 p 维随机向量 $X = (X_1, X_2, \ldots, X_p)'$ 服从参数为均值向量 μ (一个 $p \times 1$ 的列向量) 和协方差矩阵 Σ (一个 $p \times p$ 的对称正定矩阵) 的多元正态分布，如果其概率密度函数 (PDF) 为:
>
> $f(x; \mu, \Sigma) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right)$
>
> 其中:
>
> - $x = (x_1, x_2, \ldots, x_p)'$ 是随机向量 X 可能取的值。
> - $\mu = E(X)$ 是均值向量。
> - $\Sigma = D(X) = E[(X-\mu)(X-\mu)']$ 是协方差矩阵。
> - $|\Sigma|$ 是协方差矩阵 Σ 的行列式。
> - Σ^{-1} 是协方差矩阵 Σ 的逆矩阵。
> - $(x-\mu)'\Sigma^{-1}(x-\mu)$ 称为马氏距离的平方 (squared Mahalanobis distance)，它度量了点 x 到均值 μ 的距离，并考虑了各分量之间的相关性。

记号：我们通常记为 $X \sim N_p(\mu, \Sigma)$ 或 $X \sim N(\mu, \Sigma)$。下标 p 有时会省略，如果维度从上下文中是清晰的。

一元情况的对比:

当 $p=1$ 时，X 就是一个一维随机变量，μ 是其均值，Σ 就是其方差 σ^2。此时，马氏距离的平方变为 $(x-\mu)(\sigma^2)^{-1}(x-\mu)=\frac{(x-\mu)^2}{\sigma^2}$。PDF公式就退化为一元正态分布的密度函数:
$f(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

定义 2: 通过线性变换独立标准正态变量 (更一般，允许 Σ 半正定)

这种定义更为根本，因为:

> 它**不要求 Σ 必须是正定的**，从而可以包含协方差矩阵奇异的情况。

> 一个 p 维随机向量 X 服从多元正态分布，如果它可以表示为:
> $X=\mu+AY$
> 其中:
>
> - μ 是一个 $p\times 1$ 的常数向量 (即 X 的均值向量)。
> - A 是一个 $p\times q$ 的常数矩阵。
> - $Y=(Y_1,Y_2,\ldots,Y_q)'$ 是一个 q 维随机向量，其分量 Y_i 是相互独立且均服从标准正态分布 $N(0,1)$ (i.i.d. $N(0,1)$)。
>   这意味着 $E(Y)=\mathbf{0}$ (零向量) 且 $D(Y)=I_q$ (q 阶单位矩阵)。

在这种定义下:

- X 的均值向量为 $E(X)=E(\mu+AY)=\mu+AE(Y)=\mu+A\mathbf{0}=\mu$。
- X 的协方差矩阵为 $D(X)=D(\mu+AY)=D(AY)=AD(Y)A'=AI_qA'=AA'$。
  所以，$\Sigma=AA'$。

记号：$X\sim N_p(\mu,AA')$。

注:

- 这里用 $A_{p\times q}$ 来强调矩阵 A 的维度。
- 如果 Σ 是正定的，那么我们可以找到一个可逆的 $p\times p$ 矩阵 A (例如通过 Cholesky 分解 $\Sigma=LL'$，则 $A=L$) 使得 $\Sigma=AA'$。

- 如果 Σ 是半正定的但非正定 (奇异的，即 $|\Sigma| = 0$)，那么矩阵 A 的列数 q 将会小于 p ($\text{rank}(A) = \text{rank}(\Sigma) < p$)，或者 A 是一个 $p \times p$ 的奇异矩阵。这对应了数据分布在低维子空间的情况。

定义 3: 通过特征函数 (最一般)

这是**最普适的定义**，因为它不依赖于密度函数的存在性，也不直接依赖于线性变换的构造。

> 一个 p 维随机向量 X 服从多元正态分布，如果其特征函数 (characteristic function) $\phi_X(t)$ 具有以下形式：
> $\phi_X(t) = E[e^{it'X}] = \exp\left(it'\mu - \frac{1}{2}t'\Sigma t\right)$
> 其中：
>
> - $t = (t_1, t_2, \ldots, t_p)'$ 是一个 $p \times 1$ 的实数向量。
> - $i = \sqrt{-1}$ 是虚数单位。
> - $\mu = E(X)$ 是均值向量。
> - $\Sigma = D(X)$ 是协方差矩阵 (这里 Σ 可以是半正定的)。

为什么需要多种定义?

- **定义1 (PDF)**: 直观易懂，便于进行似然推断和概率计算 (当 Σ 正定时)。
- **定义2 (线性变换)**: 揭示了多元正态分布的构造本质——它是独立标准正态变量的线性组合。这个定义能够自然地处理协方差矩阵奇异的情况，并有助于理解其几何意义和性质（例如，多元正态变量的任何线性组合仍服从正态分布）。
- **定义3 (特征函数)**: 数学上最完备和普适。特征函数唯一确定一个分布，并且对于半正定的 Σ 依然良好定义。它在理论推导中非常有用，例如证明中心极限定理的多维版本。

可以证明：

> **当协方差矩阵 Σ 是正定时，这三种定义是等价的。**
>
> 当 **Σ 是半正定时，定义2和定义3仍然适用并等价，而定义1 (基于PDF) 则不适用（因为PDF不存在）**。

无论哪种定义，多元正态分布完全由其**均值向量 μ** 和**协方差矩阵 Σ** 这两个参数所确定。

> - μ 决定了分布的中心位置。

> - Σ 决定了分布的形状、离散程度以及各分量之间的相关性。对角线元素是各分量的方差，非对角线元素是协方差。

> 作者：肖宇翔
>
> 华东理工大学数学学院2022级本科生