Multivariate Normal Distirbution

Chao Cheng

July 24, 2022

For a multivariate normal distribution denoted by

$$\boldsymbol{x} \sim N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right),$$

where $\boldsymbol{x} = \left(x_1, \cdots, x_n\right)^T \in \mathcal{R}^n$, $\boldsymbol{\mu} = \left(\mu_1, \cdots, \mu_n\right)^T$ is the mean vector and

$$\boldsymbol{\Sigma} = \left(\rho_{ij}\sigma_i\sigma_j\right)_{n\times n} = \begin{pmatrix} \sigma_1^2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \vdots & \rho_{ij}\sigma_i\sigma_j & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \cdots & \sigma_n^2 \end{pmatrix}$$

is the $n \times n$ covariance matrix. Here $\sigma_1^2, \cdots, \sigma_n^2$ is the variance of $x_1, \cdots, x_n$ and ρ_{ij} is the correlation between x_i and x_j. Since it's symmetric, $\rho_{ij} = \rho_{ji}$ and $\rho_{ii} = 1$ for $i = 1, \cdots, n$.

The density function is

$$f\left(\boldsymbol{x}\right) = \left(2\pi\right)^{-n/2}\left|\boldsymbol{\Sigma}\right|^{-1/2}\exp\left(-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\right).$$

1 Conditional distribution

The inverse of a block matrix is

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} \left(A - BD^{-1}C\right)^{-1} & -\left(A - BD^{-1}C\right)^{-1}BD^{-1} \\ -D^{-1}C\left(A - BD^{-1}C\right)^{-1} & D^{-1} + D^{-1}C\left(A - BD^{-1}C\right)^{-1}BD^{-1} \end{pmatrix}$$

and

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |D| \cdot \left|A - BD^{-1}C\right|.$$

Then for

$$\begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$$

we have

$$\boldsymbol{x}_1|\boldsymbol{x}_2 \sim N\left(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}\left(\boldsymbol{x}_2 - \boldsymbol{\mu}_2\right), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right).$$

One can brute force compute the conditional density of $\boldsymbol{x}_1|\boldsymbol{x}_2$ to get this result. But since the conditional distribution of a multivariate normal is also (multivariate) normal, we can just try to figure out the mean and covariance of $\boldsymbol{x}_1|\boldsymbol{x}_2$ by introducing $\boldsymbol{z} = \boldsymbol{x}_1 + A\boldsymbol{x}_2$ where

$$A = -\Sigma_{12}\Sigma_{22}^{-1}.$$

Then we can show that

$$\begin{aligned}
&\mathrm{cov}\left(\boldsymbol{z},\boldsymbol{x}_2\right)=\mathbf{0}\\
&\mathrm{E}\left(\boldsymbol{x}_1|\boldsymbol{x}_2\right)=\mathrm{E}\left(\boldsymbol{z}-A\boldsymbol{x}_2|\boldsymbol{x}_2\right)=\boldsymbol{\mu}_1+\Sigma_{12}\Sigma_{22}^{-1}\left(\boldsymbol{x}_2-\boldsymbol{\mu}_2\right)\\
&\mathrm{Var}\left(\boldsymbol{x}_1|\boldsymbol{x}_2\right)=\mathrm{Var}\left(\boldsymbol{z}\right)=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.
\end{aligned}$$

The key point of constructing this $\boldsymbol{z}$ is that $\boldsymbol{x}_1+A\boldsymbol{x}_2$ is the residual in $\boldsymbol{x}_1$ that can not be explained by $\boldsymbol{x}_2$. Therefore it should be independent (uncorrelated actually, but we are in multivariate normal distribution scenario) of $\boldsymbol{x}_2$.

2 Bivariate normal distribution

For a bivarite normal variable (x,y) following

$$\begin{pmatrix}x\\y\end{pmatrix}\sim N\left(\begin{pmatrix}\mu_1\\\mu_2\end{pmatrix},\begin{pmatrix}\sigma_1^2&\rho\sigma_1\sigma_2\\\rho\sigma_1\sigma_2&\sigma_2^2\end{pmatrix}\right),$$

and the density function is

$$\begin{aligned}
f_{X,Y}\left(x,y\right)=&\frac{1}{2\pi}\left(\sigma_1^2\sigma_2^2-\rho^2\sigma_1^2\sigma_2^2\right)^{-1/2}\exp\left(-\frac{\begin{pmatrix}x-\mu_1&y-\mu_2\end{pmatrix}\begin{pmatrix}\sigma_2^2&-\rho\sigma_1\sigma_2\\-\rho\sigma_1\sigma_2&\sigma_1^2\end{pmatrix}\begin{pmatrix}x-\mu_1\\y-\mu_2\end{pmatrix}}{2\left(\sigma_1^2\sigma_2^2-\rho^2\sigma_1^2\sigma_2^2\right)}\right)\\
=&\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{\sigma_2^2\left(x-\mu_1\right)^2-2\rho\sigma_1\sigma_2\left(x-\mu_1\right)\left(y-\mu_2\right)+\sigma_1^2\left(y-\mu_2\right)^2}{2\sigma_1^2\sigma_2^2\left(1-\rho^2\right)}\right).
\end{aligned}$$

The probability function $F_{X,Y}\left(x,y\right)=P\left(X\le x,Y\le y\right)$ is

$$\begin{aligned}
F_{X,Y}\left(x,y\right)&=\int_{-\infty}^{x}\int_{-\infty}^{y}f_{X,Y}\left(s,t\right)\mathrm{d}s\mathrm{d}t\\
&=\int_{-\infty}^{x}\int_{-\infty}^{y}f_{Y|X}\left(s|t\right)f_X\left(t\right)\mathrm{d}s\mathrm{d}t\\
&=\int_{-\infty}^{x}f_X\left(t\right)\left(\int_{-\infty}^{y}f_{Y|X}\left(s|t\right)\mathrm{d}s\right)\mathrm{d}t.
\end{aligned}$$

For the conditional distribution $Y|X$, the density is

$$\begin{aligned}
f_{Y|X}\left(y|x\right)=&\frac{f_{X,Y}\left(x,y\right)}{f_X\left(x\right)}\\
\overset{\text{focus on } y}{\propto}&f_{X,Y}\left(x,y\right)\\
\propto&\exp\left(-\frac{\sigma_1^2y^2-2\sigma_1^2\mu_2y-2\rho\sigma_1\sigma_2\left(x-\mu_1\right)y}{2\sigma_1^2\sigma_2^2\left(1-\rho^2\right)}\right)\\
\propto&\exp\left(-\frac{y^2-2\left(\mu_2+\rho\frac{\sigma_2}{\sigma_1}\left(x-\mu_1\right)\right)y}{2\sigma_2^2\left(1-\rho^2\right)}\right),
\end{aligned}$$

which is a kernel of normal distribution hence

$$Y|X\sim N\left(\mu_2+\rho\frac{\sigma_2}{\sigma_1}\left(x-\mu_1\right),\sigma_2^2\left(1-\rho^2\right)\right)$$

Therefore

$$\int_{-\infty}^{y} f_{Y|X}(s|t)\,\mathrm{d}s = P(Y \le y|X=t) = \Phi\left(\frac{y-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(t-\mu_1)}{\sigma_2\sqrt{1-\rho^2}}\right)$$

where $\Phi(t)$ is the cdf of standard normal distribution. Therefore

$$\begin{aligned}F_{X,Y}(x,y) &= \int_{-\infty}^{x} f_X(t)\,\Phi\left(\frac{y-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(t-\mu_1)}{\sigma_2\sqrt{1-\rho^2}}\right)\mathrm{d}t \\ &= \int_{-\infty}^{x} \frac{1}{\sigma_1}\phi\left(\frac{t-\mu_1}{\sigma_1}\right)\Phi\left(\frac{y-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(t-\mu_1)}{\sigma_2\sqrt{1-\rho^2}}\right)\mathrm{d}t,\end{aligned}$$

where $\phi(t)$ is the pdf of standard normal distribution.

A R codes

```
R codes
```