Multivariate Normal Distribution I

Nipun Batra

IIT Gandhinagar

August 30, 2025

Univariate Normal Distribution

The probability density of univariate Gaussian is given as:

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

also, given as

$$f(x) \sim \mathcal{N}\left(\mu,\ \sigma^2\right)$$

with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$

Why is the denominator the way it is?

Let the normalizing constant be c and let $g(x) = e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$.

Why is the denominator the way it is?

Let the normalizing constant be c and let $g(x) = e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$.

$$1 = \int_{-\infty}^{\infty} c \cdot g(x) dx$$

Why is the denominator the way it is?

Let the normalizing constant be c and let $g(x) = e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$.

$$1 = \int_{-\infty}^{\infty} c \cdot g(x) dx$$

$$1 = \int_{-\infty}^{\infty} c e^{-\frac{(x-u)^2}{2\sigma^2}} dx$$

Why is the denominator the way it is?

Let the normalizing constant be c and let $g(x) = e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$.

$$1 = \int_{-\infty}^{\infty} c \cdot g(x) dx$$
$$1 = \int_{-\infty}^{\infty} c e^{-\frac{(x-u)^2}{2\sigma^2}} dx$$

Let's substitute $\frac{x-u}{\sqrt{2}\sigma}$ with t.

Why is the denominator the way it is?

Let the normalizing constant be c and let $g(x) = e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$.

$$1 = \int_{-\infty}^{\infty} c \cdot g(x) dx$$
$$1 = \int_{-\infty}^{\infty} c e^{-\frac{(x-u)^2}{2\sigma^2}} dx$$

Let's substitute $\frac{x-u}{\sqrt{2}\sigma}$ with t.

$$1 = \int_{-\infty}^{\infty} c e^{-t^2} dt \times \sqrt{2}\sigma$$

Why is the denominator the way it is?

Let the normalizing constant be c and let $g(x) = e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$.

$$1 = \int_{-\infty}^{\infty} c \cdot g(x) dx$$
$$1 = \int_{-\infty}^{\infty} c e^{-\frac{(x-u)^2}{2\sigma^2}} dx$$

Let's substitute $\frac{x-u}{\sqrt{2}\sigma}$ with t.

$$1 = \int_{-\infty}^{\infty} c e^{-t^2} dt \times \sqrt{2}\sigma$$
$$1 = \sqrt{2}\sigma c \times 2 \int_{0}^{\infty} e^{-t^2} dt$$

Why is the denominator the way it is?

$$\frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-t^2} dt$$

The above expression is called **error function** and is it's value is denoted by $erf(t)$. In our case, we want $erf(\infty)$ which is equal to 1.

Why is the denominator the way it is?

$$\frac{2}{\sqrt{\pi}} \int_0^\infty e^{-t^2} dt$$

The above expression is called **error function** and is it's value is denoted by $erf(t)$. In our case, we want $erf(\infty)$ which is equal to 1.

$$1 = \sqrt{2\pi}\sigma c \times \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-t^2} dt$$

Why is the denominator the way it is?

$$\frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-t^2} dt$$

The above expression is called **error function** and is it's value is denoted by $erf(t)$. In our case, we want $erf(\infty)$ which is equal to 1.

$$1 = \sqrt{2\pi}\sigma c \times \frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-t^2} dt$$
$$1 = \sqrt{2\pi}\sigma c \times 1$$

Why is the denominator the way it is?

$$\frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-t^2} dt$$

The above expression is called **error function** and is it's value is denoted by $erf(t)$. In our case, we want $erf(\infty)$ which is equal to 1.

$$1 = \sqrt{2\pi}\sigma c \times \frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-t^2} dt$$

$$1 = \sqrt{2\pi}\sigma c \times 1$$

$$\frac{1}{\sqrt{2\pi}\sigma} = c$$

Univariate Normal Distribution

Univariate Normal Distribution

Univariate Normal Distribution

Bivariate Normal Distribution

Bivariate normal distribution of two-dimensional random vector $\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\boldsymbol{\mu},\ \boldsymbol{\Sigma}\right)$$

where, mean vector $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \end{bmatrix}$

and, covariance matrix $\boldsymbol{\Sigma}$

$$\boldsymbol{\Sigma}_{i,j} := \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] = \mathrm{Cov}[X_i, X_j]$$

Bivariate Normal Distribution

Question: What is $\text{Cov}(\mathbf{X}, \mathbf{X})$?

Bivariate Normal Distribution

Question: What is $\text{Cov}(\mathbf{X}, \mathbf{X})$?
Answer: $\text{Var}(\mathbf{X}) = \text{Cov}(\mathbf{X}, \mathbf{X}) = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])]^2$
In the case of univariate normal, $\text{Var}(X)$ is written as σ^2

Bivariate Normal Distribution

Question: What is $\text{Cov}(\mathbf{X}, \mathbf{X})$?
Answer: $\text{Var}(\mathbf{X}) = \text{Cov}(\mathbf{X}, \mathbf{X}) = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])]^2$
In the case of univariate normal, $\text{Var}(X)$ is written as σ^2
Question: What is the relation between $\mathbf{\Sigma}_{i,j}$ and $\mathbf{\Sigma}_{j,i}$?

Bivariate Normal Distribution

Question: What is $\text{Cov}(\mathbf{X}, \mathbf{X})$?
Answer: $\text{Var}(\mathbf{X}) = \text{Cov}(\mathbf{X}, \mathbf{X}) = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])]^2$
In the case of univariate normal, $\text{Var}(X)$ is written as σ^2
Question: What is the relation between $\mathbf{\Sigma}_{i,j}$ and $\mathbf{\Sigma}_{j,i}$?
Answer: They are the same!

Bivariate Normal Distribution

Question: What is $\text{Cov}(\mathbf{X}, \mathbf{X})$?
Answer: $\text{Var}(\mathbf{X}) = \text{Cov}(\mathbf{X}, \mathbf{X}) = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])]^2$
In the case of univariate normal, $\text{Var}(X)$ is written as σ^2
Question: What is the relation between $\boldsymbol{\Sigma}_{i,j}$ and $\boldsymbol{\Sigma}_{j,i}$?
Answer: They are the same!
Question: What can we say about the covariance matrix $\boldsymbol{\Sigma}$?

Bivariate Normal Distribution

Question: What is $\text{Cov}(\mathbf{X}, \mathbf{X})$?
Answer: $\text{Var}(\mathbf{X}) = \text{Cov}(\mathbf{X}, \mathbf{X}) = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])]^2$
In the case of univariate normal, $\text{Var}(X)$ is written as σ^2
Question: What is the relation between $\mathbf{\Sigma}_{i,j}$ and $\mathbf{\Sigma}_{j,i}$?
Answer: They are the same!
Question: What can we say about the covariance matrix $\mathbf{\Sigma}$?
Answer: It is symmetric. Thus $\mathbf{\Sigma} = \mathbf{\Sigma}^\top$

Correlation and Covariance

If X and Y are two random variables, with means (expected values) μ_X and μ_Y and standard deviations σ_X and σ_Y, respectively, then their covariance and correlation are as follows:

$$\mathrm{Cov}_{XY} = \sigma_{XY} = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$$

Correlation and Covariance

If X and Y are two random variables, with means (expected values) μ_X and μ_Y and standard deviations σ_X and σ_Y, respectively, then their covariance and correlation are as follows:

$$\mathrm{Cov}_{XY} = \sigma_{XY} = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$$

$$\mathrm{Corr}_{XY} = \rho_{XY} = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]/(\sigma_X \sigma_Y)$$

Correlation and Covariance

If X and Y are two random variables, with means (expected values) μ_X and μ_Y and standard deviations σ_X and σ_Y, respectively, then their covariance and correlation are as follows:

$$\mathrm{Cov}_{XY} = \sigma_{XY} = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$$

$$\mathrm{Corr}_{XY} = \rho_{XY} = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]/(\sigma_X \sigma_Y)$$

so that

$$\rho_{XY} = \sigma_{XY}/(\sigma_X \sigma_Y)$$

where $\mathbb{E}$ is the expected value operator.

PDF of bivariate normal distribution

We might have seen that

$$f_{\mathbf{X}}(X_1, X_2) = \frac{\exp(\frac{-1}{2}(\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}))}{2\pi \det(\boldsymbol{\Sigma})^{\frac{1}{2}}}$$

How do we get such a weird looking formula?!

PDF of bivariate normal with no cross-correlation

Let us assume no correlation between X_1 and X_2.

We have $\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$

We have $f_{\mathbf{X}}(X_1, X_2) = f_{\mathbf{X}}(X_1) f_{\mathbf{X}}(X_2)$

$$= \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_1-\mu_1}{\sigma_1}\right)^2} \times \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_2-\mu_2}{\sigma_2}\right)^2}$$

$$= \frac{1}{\sigma_1\sigma_2 2\pi} e^{-\frac{1}{2}\left\{\left(\frac{X_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{X_2-\mu_2}{\sigma_2}\right)^2\right\}}$$

PDF of bivariate normal with no cross-correlation

Let us consider only the exponential part for now
$Q = \left(\frac{X_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{X_2-\mu_2}{\sigma_2}\right)^2$
Question: Can you write Q in the form of vectors $\mathbf{X}$ and $\boldsymbol{\mu}$?

$$= \begin{bmatrix} X_1 - \mu_1 & X_2 - \mu_2 \end{bmatrix}_{1\times 2} g(\boldsymbol{\Sigma})_{2\times 2} \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{bmatrix}_{2\times 1}$$

Here $g(\boldsymbol{\Sigma})$ is a matrix function of $\boldsymbol{\Sigma}$ that will result in σ_1^2 like terms in the denominator; also there is no cross-terms indicating zeros in right diagonal!
$g(\boldsymbol{\Sigma}) = \begin{bmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{bmatrix}_{2\times 2} = \frac{1}{\sigma_1^2\sigma_2^2} \begin{bmatrix} \sigma_2^2 & 0 \\ 0 & \sigma_1^2 \end{bmatrix}_{2\times 2} = \frac{1}{\det(\boldsymbol{\Sigma})} \operatorname{adj}(\boldsymbol{\Sigma}) = \boldsymbol{\Sigma}^{-1}$

PDF of bivariate normal with no cross-correlation

Let us consider the normalizing constant part now.

$$
\begin{aligned}
M &= \frac{1}{2\pi\ \sigma_1\ \sigma_2} \\
&= \frac{1}{2\pi \times \det(\boldsymbol{\Sigma})^{\frac{1}{2}}}
\end{aligned}
$$

Bivariate Gaussian samples with cross-correlation $\neq 0$

Bivariate Gaussian samples with cross-correlation $= 0$

Intuition for Multivariate Gaussian

Let us assume no correlation between the elements of $\mathbf{X}$. This means $\boldsymbol{\Sigma}$ is a diagonal matrix.

We have $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \sigma_n^2 \end{bmatrix}$

And,

$$\mathbb{P}(\mathbf{X}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{n}{2}} \det(\boldsymbol{\Sigma})^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})\right)$$

As seen in the case for univariate Gaussians, we can write the following for the multivariate case,

We have $f_{\mathbf{X}}(X_1, \cdots, X_n) = f_{\mathbf{X}}(X_1) \times \cdots \times f_{\mathbf{X}}(X_n)$

Intuition for Multivariate Gaussian

Now,

$$= \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_1-\mu_1}{\sigma_1}\right)^2} \times \cdots \times \frac{1}{\sigma_n\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_n-\mu_n}{\sigma_n}\right)^2}$$

$$= \frac{1}{\sigma_1 \cdots \sigma_n (2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\left\{\left(\frac{X_1-\mu_1}{\sigma_1}\right)^2 + \cdots + \left(\frac{X_n-\mu_n}{\sigma_n}\right)^2\right\}\right)$$

Taking all $\sqrt{2\pi}$ together, we get $(2\pi)^{\frac{n}{2}}$.
Similarly, taking all σ together, we get $\sigma_1 \cdots \sigma_n$. Which can be written as $\det(\mathbf{\Sigma})^{\frac{1}{2}}$, given the determinant of a digonal matrix is the multiplication of its diagonal elements.

Now, let us remove the assumption of no covariance among the elements of
X

Main idea: A correlated Gaussian is a rotated independent Gaussian[1]
Rotate input space using rotation matrix **R**.

$$\mathbb{P}(\mathbf{X}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{n}{2}} \det(\boldsymbol{\Sigma})^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{R}^\top \mathbf{X} - \mathbf{R}^\top \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{R}^\top \mathbf{X} - \mathbf{R}^\top \boldsymbol{\mu})\right)$$

$$\mathbb{P}(\mathbf{X}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{n}{2}} \det(\boldsymbol{\Sigma})^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu})^\top \mathbf{R}\boldsymbol{\Sigma}^\top (\mathbf{X} - \boldsymbol{\mu})\right)$$

[1] Neil Lawrence GPSS 2016

$$\mathbf{C} = \mathbf{R}\boldsymbol{\Sigma}^\top$$

$$\mathbb{P}(\mathbf{X}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{n}{2}} \det(\mathbf{C})^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu})^\top \mathbf{C}^{-1}(\mathbf{X} - \boldsymbol{\mu})\right)$$