Comprehensive Study Guide: The Gaussian Distribution

From First Principles to Multivariate Calculus

Tashfeen Omran

January 20, 2026

Contents

1 Phase 1: Univariate Foundations (1D)

1.1 Intuitive Construction from First Principles

The Gaussian distribution (or Normal distribution) is not an arbitrary formula; it is constructed logically from exponential decay.

1. **Base Shape:** We start with the exponential function e^x. To create a symmetric shape that decays on both sides, we square the exponent and negate it: e^{-x^2}. This creates the fundamental "bell" shape centered at 0.
2. **Shift (μ):** To center the distribution at an arbitrary point, we replace x with $(x - \mu)$.
3. **Stretch/Squeeze (σ):** To control the width (spread), we divide the input by σ.
4. **Normalization:** To ensure the total area under the curve equals 1 (a requirement for probability distributions), we divide by the constant $\sqrt{2\pi}\sigma$.

The resulting equation is:

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

1.2 The "One-Half" Factor: Why $-\frac{1}{2}$?

A common source of confusion is the factor of $\frac{1}{2}$ in the exponent. While a curve $e^{-(x/\sigma)^2}$ is still a valid bell curve, the factor $\frac{1}{2}$ is included for three specific mathematical benefits.

1.2.1 1. The Clean Derivative (Calculus Convenience)

Using the Chain Rule, the derivative of e^{-x^2} introduces a factor of 2:

$$\frac{d}{dx}e^{-x^2} = -2x \cdot e^{-x^2}$$

By including the $\frac{1}{2}$, we cancel this factor:

$$\frac{d}{dx}e^{-\frac{1}{2}x^2} = -\frac{1}{2} \cdot 2x \cdot e^{-\frac{1}{2}x^2} = -x \cdot e^{-\frac{1}{2}x^2}$$

This simplifies higher-order derivatives used in optimization and physics.

1.2.2 2. Definition of Standard Deviation

In statistics, variance is defined as the expectation of the squared deviation: $E[(x-\mu)^2]$. If we omit the $\frac{1}{2}$ from the exponent, the calculated variance of the distribution becomes $\frac{\sigma^2}{2}$. This would mean the parameter σ does not equal the standard deviation. By including the $\frac{1}{2}$, the integral evaluates such that the variance is exactly σ^2.

1.2.3 3. Geometric Inflection Points

The $\frac{1}{2}$ aligns the geometric properties of the curve with the parameter σ.

- With the $\frac{1}{2}$ factor, the **inflection points** (where the curve changes from concave down to concave up) occur exactly at $x = \mu + \sigma$ and $x = \mu - \sigma$.
- This allows for visual estimation of standard deviation simply by looking at where the "hill" stops curving downward and begins flattening out.

2 Phase 2: The Multivariate Normal Distribution

2.1 Conceptual Mapping: From Scalar to Vector

The Multivariate Normal is the generalization of the 1D bell curve to d dimensions. We replace scalar values with vectors and matrices.

Concept	**Univariate (1D)**	**Multivariate (d-dimensions)**
Variable	x (Scalar)	$\mathbf{x}$ (Vector of length d)
Center	μ (Mean scalar)	$\boldsymbol{\mu}$ (Mean Vector)
Spread	σ^2 (Variance)	Σ (Covariance Matrix, $d \times d$)
Normalization	$1/\sigma$	$\|\Sigma\|^{-1/2}$ (Inverse determinant)
Distance op.	Division	Matrix Inversion

2.2 The Formula

The Probability Density Function (PDF) for a d-dimensional vector $\mathbf{x}$ is:

$$\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

3 Phase 3: Linear Algebra Mechanics and Intuition

3.1 The "Division" Paradox: Why Σ^{-1}?

In the 1D exponent, we have $\frac{(x-\mu)^2}{\sigma^2}$. This can be rewritten as distance times inverse variance:

$$(x-\mu)\frac{1}{\sigma^2}(x-\mu)$$

In matrix algebra, **division is undefined**. We cannot write $\frac{A}{B}$. Instead, we multiply by the inverse.

$$\text{Equivalent to division by } \Sigma \implies \text{Multiplication by } \Sigma^{-1}$$

Key Intuition

Geometric Intuition: Multiplying by Σ^{-1} standardizes the space. If the distribution is an elongated, tilted ellipse (due to correlation between variables), Σ^{-1} effectively:

1. Rotates the axes to align with the ellipse.
2. Shrinks the long axis and stretches the short axis.
3. "Squishes" the ellipse back into a standard circle (or sphere).

3.2 The "Sandwich" Multiplication (Scalar Result)

The exponent of e must be a scalar (a single number). However, $\mathbf{x}$ and $\boldsymbol{\mu}$ are vectors and Σ is a matrix. We use the "Sandwich" form to resolve dimensions.

Let $d = 2$ (e.g., Height and Weight).

- $(\mathbf{x} - \boldsymbol{\mu})$ is a row vector (1×2).
- Σ^{-1} is a square matrix (2×2).
- $(\mathbf{x} - \boldsymbol{\mu})^T$ is a column vector (2×1).

The multiplication proceeds as:

$$\underbrace{(1 \times 2)}_{\text{Row}} \cdot \underbrace{(2 \times 2)}_{\text{Matrix}} \cdot \underbrace{(2 \times 1)}_{\text{Column}}$$

$$\underbrace{(1 \times 2)}_{\text{Result is Row}} \cdot (2 \times 1) \implies \mathbf{1 \times 1} \text{ (Scalar)}$$

This scalar result represents the **Mahalanobis Distance**: the squared distance of point $\mathbf{x}$ from the mean $\boldsymbol{\mu}$, corrected for the "shape" (covariance) of the distribution.

3.3 Dimensionality and Components

- **Length d Vector:** If we are modeling 3 features (e.g., x, y, z), then $d = 3$. The mean $\boldsymbol{\mu}$ is a list of 3 numbers.
- **Covariance Matrix (Σ):** This is always $d \times d$. For $d = 2$:

$$\Sigma = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix}$$

  - **Diagonal (σ_{xx}, σ_{yy}):** Variances of individual variables (width of the hill along axes).
  - **Off-Diagonal (σ_{xy}):** Covariance. If non-zero, the variables are correlated, and the "hill" appears tilted/rotated when viewed from above.