

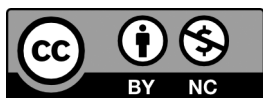
Pedagogical Reflection on the Multivariate Normal Distribution

Shaowen Yu Xiangqi Zheng*

School of Mathematics, East China University of Science and Technology, Shanghai

Abstract: This paper proposes a novel pedagogical approach to teaching the multivariate normal by using an equivalent definition that demands less algebraic and analytical sophistication than the conventional formulation, while more effectively emphasizing its distinctive properties. To further enhance student engagement, the approach integrates a case study on AI-based image processing as an application of the multivariate normal distribution within the instructional framework.

Keywords: Multivariate normal; Equivalent definition; Image processing; Pedagogical reflection



Copyright © 2025 by author (s) and SciScan Publishing Limited

This article is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/). <https://creativecommons.org/licenses/by-nc/4.0/>

1 Introduction and Motivation

The normal distribution is arguably the most important in probability theory, and its multivariate generalization is even more prominent in multivariate statistics. From a theoretical standpoint, multivariate central limit theorems ensure that a wide class of statistics is approximately normal, providing distributional foundations for confidence regions, hypothesis tests, and likelihood-based inference in high dimensions. The multivariate normal's algebraic conveniences — closed forms for linear images and sums of independent Gaussians; block-matrix formulas for conditioning; and explicit relationships between eigenstructure and principal components — make it a natural backbone for methods such as linear/quadratic discriminant analysis (LDA/QDA), and Gaussian linear models (Murphy, 2012). In multiple testing and reproducibility, jointly normal models under weak dependence enable asymptotic analyses of family-wise error rate (FWER), yielding explicit correction terms relative to independence via multivariate normal tail calculus in nearly-independent regimes (Das & Bhandari, 2025).

Beyond statistics, the (multivariate) normal family is a practical modeling workhorse across engineering and operations. In operations management, newsvendor-type inventory models frequently assume (or approximate) demand by a normal law, enabling closed-form critical-fractile rules, tractable price-demand coupling, and sensitivity analyses; recent work with price-dependent normal demand and partial backorders shows how normality supports both analytic

Research/Funding project: This work was supported by the Higher Education Scientific Research Planning Project (Grant Nos. 23SX0419 and 24SX0404).

*Corresponding author: Xiangqi Zheng, School of Mathematics, East China University of Science and Technology, Associate Professor. Research direction: Probability.

Article Citation: Yu, S. W., & Zheng, X. Q. (2025). Pedagogical Reflection on the Multivariate Normal Distribution. *Guide to Education Innovation*, 5(4), 182-187.

solutions and robust managerial insights (Pando et al., 2024). In geotechnical engineering, multivariate uncertainty in geo-material parameters is often represented to support reliability-based design (Han et al., 2023). In civil/transportation infrastructure, ideas from the multivariate normal inform constitutive modeling: for example, Monte-Carlo fitting of multi-segment bond-slip laws at shield-tunnel interfaces leverages multivariate normal structure and confidence statistics to quantify turning points, stiffness stages, and bond strength (Wang & Jin, 2025). In process safety, normal-distribution methodology helps analyze dispersed dust mass and its impact on minimum ignition temperature in furnace tests, linking dispersion pressure to ignition behavior and flame propagation speeds — illustrating how Gaussian modeling informs hazard parameters (Guo et al., 2025).

In most Chinese probability textbooks, there are two approaches to introducing the multivariate normal distribution: an algebraic (Definition A) approach that presents its probability density function using matrix notation, and an analytic approach (Definition B) that defines it via the characteristic function.

Definition A (Algebraic, via density): Let $\mu \in R^n$ and $\Sigma \in R^{n \times n}$ be symmetric and positive definite. A random vector $X \in R^n$ is multivariate normal with mean μ and covariance Σ if it has Lebesgue density

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

where $|\Sigma|$ denotes the determinant of Σ , Σ^{-1} its inverse.

Definition B (Analytic, via characteristic function): Let $\mu \in R^n$ and $\Sigma \in R^{n \times n}$ be symmetric and positive semidefinite. A random vector X is multivariate normal with mean μ and covariance Σ , if and only if its characteristic function is

$$\varphi_X(t) = E\left[e^{it^T X}\right] = \exp(it^T \mu - \frac{1}{2}t^T \Sigma t), \quad t \in R^n.$$

Definition A presupposes a certain level of algebraic maturity, whereas Definition B can only be introduced after the material on characteristic functions has been covered. In probability courses designed for engineering students, characteristic functions are typically omitted; even among mathematics majors, not all students master this tool on first exposure. Perhaps because both approaches place relatively high demands on students' mathematical background, most textbooks — despite the wide practical use of the multivariate normal across disciplines — still treat it as asterisked optional material. In our teaching experience, students tend either to skip such topics or to engage with them only superficially, which in turn creates difficulties later in mathematical statistics and related applications. To lower the barrier to entry, we recast the topic by taking an equivalent property of the multivariate normal as the operative definition, anchored in the well-known univariate normal. And at last, we illustrate how the multivariate normal is used in AI-driven image processing (computer vision). This design has improved engagement in our large-enrollment classes and facilitated a smoother transition to subsequent statistical modeling.

2 Operational Definition and its Justification

As the most famous continuous distribution, the one-dimensional normal has long occupied a central place in the teaching of probability; students — whether majoring in mathematics or in science and engineering — generally master it well. The bivariate normal is likewise a key component of the chapter on random vectors and serves as a classic vehicle for introducing covariance, correlation, and related concepts. In our instructional practice, immediately

after covering the bivariate case, we introduce the multivariate normal as a direct extension and adopt the following definition.

Definition C (Structural, via linear combinations): A random vector X is multivariate normal if for any $\alpha \in R^n$, the linear combination $\alpha^T X$ is (univariate) normal; when $\text{Var}(\alpha^T X) = 0$ for some α , we regard $\alpha^T X$ as a degenerate normal at its mean.

This definition circumvents the formally intricate joint-density formulation and the more demanding notion of characteristic functions, while leveraging students' familiarity and comfort with the univariate normal. A common beginner's misconception is that "if each component is normally distributed, then the vector is multivariate normal"; framing the concept this way naturally dispels that misunderstanding and simultaneously highlights the distinctive features of the multivariate normal. We now justify the validity of this definition.

Theorem: For any random vector $X \in R^n$, the following two properties are equivalent.

(1) The characteristic function of X is

$$\varphi_X(t) = E[e^{it^T X}] = \exp(it^T \mu - \frac{1}{2} t^T \Sigma t), \quad t \in R^n;$$

(2) For any $\alpha \in R^n$ the linear combination $\alpha^T X$ is (univariate) normal.

Proof. (1) $\Rightarrow$ (2) is obvious. So we only prove (2) $\Rightarrow$ (1). If for any $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)^T$, $\alpha^T X = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n$ is normal, then the characteristic function of $\alpha^T X$ is

$$\varphi_{\alpha^T X}(t) = E(e^{it\alpha^T X}) = \exp\left\{itE(\alpha^T X) - \frac{t^2}{2}\text{Var}(\alpha^T X)\right\} \quad (1)$$

where $E(\alpha^T X)$ is the expectation of $\alpha^T X$ and $\text{Var}(\alpha^T X)$ is the variance. Let $\theta_i = EX_i, i = 1, 2, \dots, n$.

By the linearity of mathematical expectation, we have

$$E(\alpha^T X) = \alpha_1 \theta_1 + \alpha_2 \theta_2 + \dots + \alpha_n \theta_n = \alpha^T \theta \quad (2)$$

Let the covariance matrix of the random vector $X = (X_1, X_2, \dots, X_n)^T$ be $C = [C_{ij}]_{n \times n}$, where $C_{ij} = \text{cov}(X_i, X_j)$, $i, j = 1, 2, \dots, n$. By standard properties of covariance matrices, C is an $n \times n$ symmetric positive semidefinite matrix, and moreover,

$$\text{Var}(\alpha^T X) = \alpha^T C \alpha \quad (3)$$

Substituting (2) and (3) into (1) yields that, $\varphi_{\alpha^T X}(t) = \exp\left\{it\alpha^T \theta - \frac{t^2}{2}\alpha^T C \alpha\right\}$. Thus, for any $\eta = (\eta_1, \eta_2, \dots, \eta_n)^T$, the characteristic function of $X = (X_1, X_2, \dots, X_n)$ is

$$\varphi_X(\eta_1, \eta_2, \dots, \eta_n) = E(e^{i\eta^T X}) = \varphi_{\eta^T X}(1) = \exp\left\{i\eta^T \theta - \frac{1}{2}\eta^T C \eta\right\}.$$

That completes the proof.

We have thus completed the justification of the definition. For public probability courses aimed at non-mathematics majors in science and engineering, one may introduce the multivariate normal directly via this definition and omit the proof. In courses for mathematics majors, after using this definition as the entry point, the above argument can be followed to derive the characteristic function of the multivariate normal.

Moreover, Definition C makes several properties of the multivariate normal essentially immediate.

Property 1: Any subvector of a multivariate normal random vector is again multivariate (or univariate) normal.

Property 2: If $X = (X_1, X_2, \dots, X_n)^T$ and $Y = (Y_1, Y_2, \dots, Y_m)^T$ are independently multivariate normal, then the concatenated random vector $(X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m)^T$ is also multivariate normal.

3 A Classroom Case: Image Blurring

As a foundational operation in computer vision, image blurring plays a role in modern AI that far exceeds traditional “preprocessing”. Many computer-vision pipelines incorporate blurring as a core stage, with applications extending beyond simple noise suppression to model training, data augmentation, and privacy protection. As a key mathematical tool for blurring, the multivariate normal provides a rigorous probabilistic framework and flexible control mechanisms — for example, classical Gaussian blur, anisotropic Gaussian blur, Mahalanobis distance–adaptive blur, and Gaussian–Markov random field (GMRF)–based blurring, among others. We illustrate the use of the multivariate normal via the classical Gaussian blur.

Digital raster images are typically stored as two-dimensional matrices of pixels, with each pixel location assigned one or more values encoding color information; together with associated metadata, these form a complete image file. We consider the following simple schematic as an example (Figure 1).



Figure 1 Example of a Simple Pixel Image

Figure 1 can be viewed as a 3×3 pixel matrix. Without loss of generality, let the coordinate matrix of the center and its neighboring points be

$$\begin{pmatrix} (-1,1) & (0,1) & (1,1) \\ (-1,0) & (0,0) & (1,0) \\ (-1,-1) & (0,-1) & (1,-1) \end{pmatrix}.$$

It is obvious that the pixel matrix can be given by

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Intuitively, the essence of image blurring is to let each pixel’s value “blend” with those of its neighbors — much like pigment diffusing on paper: each tiny color patch (pixel) “bleeds” slightly outward, and the final color equals:

$$\text{its own color} \times (\text{its weight}) + \text{neighbors' colors} \times (\text{their respective weights}).$$

When all pixels diffuse simultaneously, previously sharp edges become smooth. It is natural, therefore, that closer neighbors should carry greater influence (“a louder voice”) in determining the final color. The joint probability density function of the bivariate normal fits this requirement precisely, as illustrated in Figure 2.

Bivariate Normal Density Function

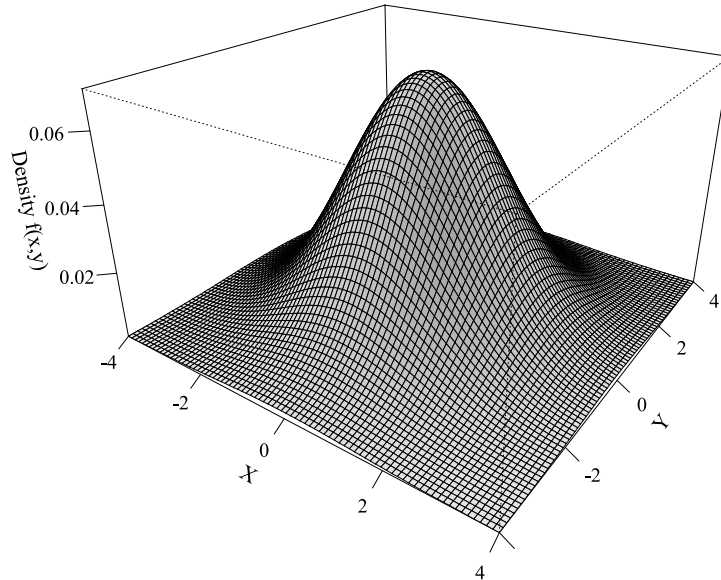


Figure 2 Density of 2-Dimensional Normal

Note that, although the joint density of a bivariate normal is defined on the entire plane, by the 3σ -rule, the influence of points more than 3σ away from the center is negligible. Consequently, we need not work over the whole plane; in practice, one truncates to a finite window — and, if desired, to an even smaller region. In this example, we set the blur radius to 1 and restrict our attention to points strictly within a distance of 1 of the center. For simplicity, we specify the parameters of the bivariate normal $N(\theta, C)$ as follows:

$$\theta = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, C = \begin{pmatrix} 2.25 & 0 \\ 0 & 2.25 \end{pmatrix}.$$

Substituting the coordinates of each point, we obtain the values of the bivariate normal density at the center and at the eight surrounding points at a distance of 1 (i.e., over the 3×3 neighborhood):

$$\begin{pmatrix} 0.045 & 0.057 & 0.045 \\ 0.057 & 0.071 & 0.057 \\ 0.045 & 0.057 & 0.045 \end{pmatrix}.$$

After normalization, we obtain the blur weight matrix with radius 1 as follows:

$$\begin{pmatrix} 0.094 & 0.119 & 0.094 \\ 0.119 & 0.148 & 0.119 \\ 0.094 & 0.119 & 0.094 \end{pmatrix}.$$

Using this weight matrix, we take a weighted average for the center pixel (0,0) and get

$$0 \times 0.094 + 1 \times 0.119 + 0 \times 0.094 + 1 \times 0.148 + 0 \times 0.119 + 0 \times 0.094 + 1 \times 0.119 + 0 \times 0.094 = 0.386.$$

Thus, the blurred center value is 0.386, clearly reflecting the influence of the surrounding white squares. Repeating

the same procedure with each pixel as the center — computing its weight matrix from the bivariate normal density and then taking the corresponding weighted average — yields the blurred value for every pixel, thereby completing the blurring of the entire image.

4 Conclusion and Teaching Recommendations

This paper offers a set of pedagogical reflections on teaching the multivariate normal. In practice, starting from the linear-combination characterization (Definition C) measurably lowers the entry barrier, and the image-blurring case study reliably boosts student engagement. As a concrete sequence, we recommend opening with a concise recap of the univariate and bivariate normal, then introducing the alternative definition of the multivariate normal. For the application segment, first implement Gaussian blurring with the parameter settings given in the text and show the resulting image; next, vary the parameters to display contrasting blur patterns and prompt students to interpret the changes. And the suggested assessments are: (1) Use the new definition to derive the two properties listed in the paper; (2) Implement image blurring in code under multiple parameter choices and report their observations in small groups.

References

- [1] Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. Cambridge: MIT Press.
- [2] Das, N., & Bhandari, S. K. (2025). FWER for normal distribution in nearly independent setup. *Statistics & Probability Letters*, 219, 110340.
- [3] Pando, V., San-José, L. A., Sicilia, J., & Alcaide-López-de-Pablo, D. (2024). Pricing decision in a newsvendor model with partial backorders under a normal probability distribution for the demand. *Applied Mathematical Modelling*, 132, 57–72.
- [4] Han, L., Liu, H., Zhang, W., & Wang, L. (2023). A comprehensive comparison of copula models and multivariate normal distribution for geo-material parametric data. *Computers and Geotechnics*, 164, 105777.
- [5] Wang, Z., & Jin, H. (2025). Study on the constitutive model of invert-filling–shield tunnel interface based on multivariate normal distribution. *Engineering Fracture Mechanics*, 316, 110878.
- [6] Guo, G. S., Cheng, Y. C., Liao, S. W., & Shu, C. M. (2025). A normal distribution-based approach to evaluate the effect of dispersion pressure on the minimum ignition temperature of dust clouds. *Advanced Powder Technology*, 36, 104763.