Multivariate Normal Distribution

Statistics · Module 8 — Multivariate Linear Models

Node 8.2

Position in Knowledge Graph

System: Statistics. Structural Domain: Module 8 — Multivariate and Linear Models. Node 8.2 — Multivariate Normal Distribution.

The multivariate normal (MVN) is the canonical multivariate distribution, completely determined by its first two moments. It is the reference distribution for all finite-sample inference in linear models. This node depends on the covariance structure (Node 8.1) and distribution theory (Node 1.4), and enables linear models (Node 8.3), inference (Node 8.5), and PCA (Node 8.7).

Formal Definition

A random vector $X \in \mathbb{R}^p$ has the multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$ if every linear combination is univariate normal:

$$a^\top X \sim \mathcal{N}(a^\top \mu,\ a^\top \Sigma a) \quad \forall\, a \in \mathbb{R}^p.$$

When $\Sigma \succ 0$, the density is

$$f(x) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\!\left(-\tfrac{1}{2}(x-\mu)^\top \Sigma^{-1} (x-\mu)\right).$$

The moment generating function is $M(t) = \exp(t^\top \mu + \frac{1}{2} t^\top \Sigma t)$.

Intuition

The MVN generalizes the bell curve to multiple dimensions. The density is constant on ellipsoids centered at μ, with axes aligned along the eigenvectors of Σ and half-lengths proportional to the square roots of the eigenvalues. The key structural property is that the entire dependence structure is captured by Σ: for jointly normal vectors, uncorrelatedness implies independence, and all conditional and marginal distributions remain normal. This makes the MVN the only distribution where second-moment analysis is sufficient for complete probabilistic characterization.

Structural Role

The MVN is the reference distribution for multivariate statistics. Its closure under affine transformations, the normality of marginals and conditionals, and the characterization of quadratic forms underlie all finite-sample distributional results in the normal linear model — t-tests, F-tests, confidence ellipsoids. The Wishart distribution (the sampling distribution of the sample covariance) is derived from the MVN. In asymptotic theory, the MVN appears as the limit distribution in the multivariate CLT.

Mathematical Development

Affine closure. If $X \sim \mathcal{N}(\mu, \Sigma)$ and $Y = AX + b$ with $A \in \mathbb{R}^{q \times p}$, then

$$Y \sim \mathcal{N}(A\mu + b,\ A\Sigma A^\top).$$

Proof. For any $c \in \mathbb{R}^q$, $c^\top Y = c^\top AX + c^\top b = (A^\top c)^\top X + c^\top b$, which is univariate normal since X is MVN. The mean is $c^\top(A\mu + b)$ and variance is $c^\top A\Sigma A^\top c$. $\square$

Marginal distributions. Partition $X = (X_1^\top, X_2^\top)^\top$ with $X_1 \in \mathbb{R}^{p_1}$, $X_2 \in \mathbb{R}^{p_2}$, and correspondingly

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Theorem. $X_1 \sim \mathcal{N}(\mu_1, \Sigma_{11})$.

Proof. $X_1 = (I_{p_1}, 0)X$, so by affine closure, $X_1 \sim \mathcal{N}(\mu_1, \Sigma_{11})$. $\square$

Conditional distributions. The conditional distribution of $X_1 \mid X_2 = x_2$ is

$$X_1 \mid X_2 = x_2 \ \sim\ \mathcal{N}\left(\mu_{1|2},\ \Sigma_{1|2}\right),$$

where

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \qquad \Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Proof. Define $Z = X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2$. Then $\operatorname{Cov}(Z, X_2) = \Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22} = 0$. Since (Z, X_2) is jointly normal (as an affine function of X) and uncorrelated, $Z \perp\!\!\!\perp X_2$. Therefore $X_1 \mid X_2 = x_2$ has the same distribution as $Z + \Sigma_{12}\Sigma_{22}^{-1}x_2$, giving the stated mean and variance. $\square$

The conditional variance $\Sigma_{1|2}$ is the Schur complement of Σ_{22} in Σ. It does not depend on the conditioning value x_2 — this is the homoskedasticity of the MVN conditional.

Independence and uncorrelatedness. For jointly normal X_1, X_2:

$$\Sigma_{12} = 0 \iff X_1 \perp\!\!\!\perp X_2.$$

The forward direction follows because uncorrelated jointly normal vectors have independent MGFs: $M_{X_1,X_2}(t_1, t_2) = M_{X_1}(t_1)M_{X_2}(t_2)$.

Quadratic forms. If $X \sim \mathcal{N}(\mu, \Sigma)$ with $\Sigma \succ 0$, then

$$(X - \mu)^\top \Sigma^{-1}(X - \mu) \sim \chi^2_p.$$

Proof. Let $Z = \Sigma^{-1/2}(X - \mu) \sim \mathcal{N}(0, I_p)$. Then $(X - \mu)^\top\Sigma^{-1}(X - \mu) = Z^\top Z = \sum_{j=1}^p Z_j^2 \sim \chi^2_p$. $\square$

More generally, if $X \sim \mathcal{N}(0, I_n)$ and A is symmetric idempotent with rank r, then $X^\top AX \sim \chi^2_r$. This is the fundamental result connecting projection matrices to chi-squared distributions (used in Node 8.5).

Cochran's theorem. Let $X \sim \mathcal{N}(0, \sigma^2 I_n)$ and let $A_1, \ldots, A_m$ be symmetric matrices with $\sum_j A_j = I_n$ and $\operatorname{rank}(A_j) = r_j$ with $\sum_j r_j = n$. If each A_j is idempotent (equivalently, each $A_j \succeq 0$ and the rank condition holds), then $X^\top A_1 X/\sigma^2, \ldots, X^\top A_m X/\sigma^2$ are independent $\chi^2_{r_j}$ random variables.

Proof sketch. Since A_j is symmetric idempotent with rank r_j, it has eigendecomposition $A_j = U_j U_j^\top$ where $U_j \in \mathbb{R}^{n \times r_j}$ has orthonormal columns. The condition $\sum A_j = I_n$ implies the columns of $U_1, \ldots, U_m$ form an orthonormal basis of $\mathbb{R}^n$. Define $Z_j = U_j^\top X/\sigma \sim \mathcal{N}(0, I_{r_j})$. Then $X^\top A_j X/\sigma^2 = Z_j^\top Z_j \sim \chi^2_{r_j}$, and the Z_j are independent because $U_j^\top U_k = 0$ for $j \neq k$. □

Characterization via MGF. A random vector X is MVN if and only if its MGF has the form $M(t) = \exp(t^\top \mu + t^\top \Sigma t/2)$. This characterization is useful for proving closure properties: if X has this MGF and $Y = AX + b$, then $M_Y(t) = \exp(t^\top (A\mu + b) + t^\top A \Sigma A^\top t/2)$, confirming $Y \sim \mathcal{N}(A\mu + b, A\Sigma A^\top)$.

Singular MVN. When Σ is positive semidefinite but not positive definite (rank $r < p$), X is supported on an r-dimensional affine subspace and has no density with respect to Lebesgue measure on $\mathbb{R}^p$. The distribution is still well-defined: $X = \mu + BZ$ where $Z \sim \mathcal{N}(0, I_r)$ and $B \in \mathbb{R}^{p \times r}$ satisfies $BB^\top = \Sigma$. All linear combination properties, marginal/conditional formulas, and the Portmanteau characterization remain valid.

Stein's lemma. If $X \sim \mathcal{N}(\mu, \Sigma)$ and $g : \mathbb{R}^p \to \mathbb{R}$ is differentiable with $\mathbb{E}[|\nabla g(X)|] < \infty$, then

$$\operatorname{Cov}(X_j, g(X)) = \sum_{k=1}^{p} \Sigma_{jk} \mathbb{E}\left[\frac{\partial g}{\partial x_k}(X)\right].$$

This identity characterizes the normal distribution and is the basis for Stein's method for proving normal approximation. In the scalar case: $\operatorname{Cov}(X, g(X)) = \sigma^2 \mathbb{E}[g'(X)]$.

Multivariate CLT. If $X_1, \ldots, X_n$ are i.i.d. in $\mathbb{R}^p$ with $\mathbb{E}[X_i] = \mu$ and $\operatorname{Cov}(X_i) = \Sigma$, then $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \Sigma)$. The limit is MVN regardless of the distribution of X_i — the Gaussian character is inherited from the CLT, not assumed. This is the reason the MVN appears throughout asymptotic statistics.

Wishart distribution. Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma)$. The matrix $W = \sum_{i=1}^n X_i X_i^\top$ follows the Wishart distribution $W_p(n, \Sigma)$. The density (for $n \geq p$) is

$$f(W) = \frac{|W|^{(n-p-1)/2}}{2^{np/2} |\Sigma|^{n/2} \Gamma_p(n/2)} \exp\left(-\tfrac{1}{2} \operatorname{tr}(\Sigma^{-1} W)\right),$$

where Γ_p is the multivariate gamma function. The Wishart is the multivariate generalization of the chi-squared distribution: when $p = 1$, $W_1(n, \sigma^2) = \sigma^2 \chi^2_n$.

Worked Examples

Example 1. Let $X \sim \mathcal{N}\left(\begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}4 & 2\\2 & 3\end{pmatrix}\right)$. The conditional distribution of $X_1 \mid X_2 = 5$ is $\mathcal{N}(1 + (2/3)(5 - 2),\ 4 - 4/3) = \mathcal{N}(3,\ 8/3)$. The conditional mean is a linear function of the conditioning value: a unit increase in X_2 shifts the conditional mean of X_1 by $\Sigma_{12}/\Sigma_{22} = 2/3$. The conditional variance $8/3$ is less than the marginal variance 4, reflecting the information gained from observing X_2.

Example 2. Let $Z \sim \mathcal{N}(0, I_3)$. The quadratic form $Q = Z_1^2 + Z_2^2 + Z_3^2 \sim \chi^2_3$. Let H be the projection onto $\operatorname{span}(1,1,1)^\top/\sqrt{3}$, so $H = \mathbf{1}\mathbf{1}^\top/3$. Then $Z^\top H Z/1 = (\bar{Z})^2 \cdot 3 \sim \chi^2_1$ and $Z^\top (I - H) Z \sim \chi^2_2$, with the two quadratic forms independent by Cochran's theorem. This decomposition is the simplest instance of the ANOVA F-test structure.

Example 3 (Conditional distribution computation). Let $X \sim \mathcal{N}\left(0, \begin{pmatrix} 1 & 0.5 & 0.3 \\ 0.5 & 1 & 0.4 \\ 0.3 & 0.4 & 1 \end{pmatrix}\right)$. Partition as $X_1 = X_1$ (scalar) and $X_2 = (X_2, X_3)^\top$. Then $\Sigma_{12} = (0.5, 0.3)$, $\Sigma_{22} = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$, $\Sigma_{22}^{-1} = \frac{1}{0.84}\begin{pmatrix} 1 & -0.4 \\ -0.4 & 1 \end{pmatrix}$. The conditional mean of $X_1 \mid X_2 = x_2, X_3 = x_3$ is $\Sigma_{12}\Sigma_{22}^{-1}(x_2, x_3)^\top = \frac{1}{0.84}(0.5 - 0.12)x_2 + (0.3 - 0.2)x_3) = \frac{1}{0.84}(0.38x_2 + 0.1x_3)$. The conditional variance is $1 - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = 1 - \frac{1}{0.84}(0.5 \cdot 0.38 + 0.3 \cdot 0.1) \approx 0.738$.

Cross-Links

AI. Gaussian processes use the MVN as the finite-dimensional distribution of function values; the conditional distribution formula gives the posterior predictive distribution. Variational autoencoders use the MVN as the latent space prior and the reparametrization trick exploits affine closure.

Economics. The MVN is the standard assumption in portfolio theory (Markowitz) and CAPM; the mean-variance efficient frontier is derived from the MVN assumption on asset returns.

Linear Algebra. The MVN density is an exponential of a negative definite quadratic form; level sets are ellipsoids determined by the spectral decomposition of Σ^{-1}. The Schur complement in the conditional variance formula is the same Schur complement that appears in block matrix inversion.

Quantum Mechanics. Gaussian quantum states are characterized by the first two moments of the quadrature operators, with the covariance matrix subject to the Heisenberg uncertainty constraint $\Sigma + i\Omega/2 \succeq 0$. The conditional distribution formula has a direct quantum analogue in Gaussian state conditioning (homodyne detection).

Structural Compression

Invariant. The MVN is determined by its first two moments; all higher-order cumulants vanish. Marginals, conditionals, and affine transformations of MVN vectors are MVN.

Mechanism. The density is an exponential of a negative definite quadratic form in x; level sets are ellipsoids aligned with the eigenvectors of Σ^{-1}. The MGF $\exp(t^\top \mu + t^\top \Sigma t/2)$ encodes all moments and makes closure under affine maps immediate.

Cross-System Echo. The MVN is the maximum-entropy distribution with fixed mean and covariance (information-theoretic characterization). In information geometry, the Gaussian family forms a statistical manifold with Fisher metric determined by the precision matrix. The Wishart distribution generalizes the chi-squared to matrix-valued sufficient statistics.

Exercises

1. Let $X \sim \mathcal{N}(\mu, \Sigma)$ with $\Sigma \succ 0$. Derive the density by applying the change-of-variables formula to $Z = \Sigma^{-1/2}(X - \mu) \sim \mathcal{N}(0, I_p)$.
2. Prove that if $X \sim \mathcal{N}(\mu, \Sigma)$ and $\Sigma_{12} = 0$, then $X_1 \perp\!\!\!\perp X_2$, using the MGF factorization.
3. Let $X \sim \mathcal{N}(0, \Sigma)$ with $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. Compute the conditional distribution $X_1 \mid X_2 = x_2$ and show that the regression of X_1 on X_2 has slope ρ and residual variance $1 - \rho^2$.
4. Prove Cochran's theorem for the case $m = 2$: if $A_1 + A_2 = I_n$ with A_1, A_2 symmetric and idempotent, then $X^\top A_1 X$ and $X^\top A_2 X$ are independent chi-squared.

5. Let $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \Sigma)$. Show that $\bar{X}$ and $S = (n-1)^{-1} \sum (X_i - \bar{X})(X_i - \bar{X})^\top$ are independent, and that $(n-1)S \sim W_p(n-1, \Sigma)$.
6. Compute the entropy of $\mathcal{N}(\mu, \Sigma)$ and show it depends only on $|\Sigma|$, confirming the maximum-entropy characterization.
7. Argue, structurally, why the equivalence of uncorrelatedness and independence under joint normality is a special property of the Gaussian and fails for general distributions, connecting this to the role of higher-order cumulants.

Statistics · Module 8 — Multivariate Linear Models · Node 8.2
Concepts: random-variables, variance-and-covariance, inner-product
Prerequisites: 8.1, 1.4
Enables: 8.3, 8.5, 8.7

hbar.university — own.your.cognition