4 The Multivariate Normal Distribution

The following are three possible definitions of the multivariate normal distribution (MVN). Given a vector $\boldsymbol{\mu}$ and a positive semidefinite matrix $\boldsymbol{\Sigma}$, $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if:

4.1 Definition: For a positive definite $\boldsymbol{\Sigma}$, the density function of $\mathbf{Y}$ is

$$f_{\mathbf{Y}}(\mathbf{y}) = (2\pi)^{-n/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\}.$$

4.2 Definition: The moment generating function (m.g.f.) of $\mathbf{Y}$ is

$$M_{\mathbf{Y}}(\mathbf{t}) \equiv E[e^{\mathbf{t}'\mathbf{Y}}] = \exp\{\boldsymbol{\mu}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\}.$$

4.3 Definition: $\mathbf{Y}$ has the same distribution as $\mathbf{AZ} + \boldsymbol{\mu}$, where $\mathbf{Z} = (Z_1, \ldots, Z_k)$ is a random sample from $N(0,1)$ and $\mathbf{A}_{n\times k}$ satisfies $\mathbf{AA}' = \boldsymbol{\Sigma}$.

4.4 Theorem: Definitions 4.1, 4.2, and 4.3 are equivalent for $\boldsymbol{\Sigma} > 0$ (positive definite). Definitions 4.2 and 4.3 are equivalent for for $\boldsymbol{\Sigma} \geq 0$ (positive semidefinite). If $\boldsymbol{\Sigma}$ is not positive definite, then $\mathbf{Y}$ has a singular MVN distribution and no density function exists.

4.5 Theorem: If $\mathbf{Z} = (Z_1, \ldots, Z_n)$ is a random sample from $N(0,1)$, then $\mathbf{Z}$ has the $N(\mathbf{0_n}, \mathbf{I_{n\times n}})$ distribution.

4.6 Theorem: $E[\mathbf{Y}] = \boldsymbol{\mu}$, $\mathrm{cov}(\mathbf{Y}) = \boldsymbol{\Sigma}$.

4.7 Example: Let $\mathbf{Z} = (Z_1, Z_2)' \sim N_2(\mathbf{0}, \mathbf{I})$, and let $\mathbf{A}$ be the linear transformation matrix

$$\mathbf{A} = \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix}.$$

Let $\mathbf{Y} = (Y_1, Y_2)'$ be the linear transformation

$$\mathbf{Y} = \mathbf{AZ} = \begin{pmatrix} (Z_1 - Z_2)/2 \\ (Z_2 - Z_1)/2 \end{pmatrix}.$$

By Definition 4.3 $\mathbf{Y} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = \mathbf{AA}'$.

4.8 Theorem: If $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{C}_{p\times n}$ is a constant matrix of rank p, then $\mathbf{CY} \sim N_p(\mathbf{C}\boldsymbol{\mu}, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}')$.

4.9 Theorem: $\mathbf{Y}$ is MVN if and only if $\mathbf{a}'\mathbf{Y}$ is normally distributed for all non-zero constant vectors $\mathbf{a}$.

4.10 Theorem: Let $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I})$, and let $\mathbf{T}$ be an orthogonal constant matrix. Then $\mathbf{TY} \sim N_n(\mathbf{T}\boldsymbol{\mu}, \sigma^2\mathbf{I})$.

4.11 Note: Theorem 4.10 says that mutually independent normal random variables with common variance remain mutually independent with common variance under orthogonal transformations. Orthogonal matrices correspond to rotations and reflections about the origin, i.e., they preserve the vector length:

$$||\mathbf{T}\mathbf{y}||^2 = (\mathbf{T}\mathbf{y})'(\mathbf{T}\mathbf{y}) = \mathbf{y}'\mathbf{T}'\mathbf{T}\mathbf{y} = \mathbf{y}'\mathbf{y} = ||\mathbf{y}||^2.$$

Let $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be partitioned as

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix},$$

where $\mathbf{Y}_1$ is $p \times 1$ and $\mathbf{Y}_2$ is $q \times 1$, $(p + q = n)$. The mean and covariance matrix are correspondingly partitioned as

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} = \begin{pmatrix} \operatorname{cov}(\mathbf{Y}_1) & \operatorname{cov}(\mathbf{Y}_1, \mathbf{Y}_2) \\ \operatorname{cov}(\mathbf{Y}_2, \mathbf{Y}_1) & \operatorname{cov}(\mathbf{Y}_2) \end{pmatrix}.$$

4.12 Theorem: The marginal distributions are $\mathbf{Y}_1 \sim N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $\mathbf{Y}_2 \sim N_q(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$.

4.13 Theorem: Uncorrelated implies independent: $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are independent if and only if $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}' = \mathbf{0}$.

4.14 Theorem: If $\boldsymbol{\Sigma}$ is positive definite, then the conditional distribution of $\mathbf{Y}_1$ given $\mathbf{Y}_2$ is

$$\mathbf{Y}_1 | \mathbf{Y}_2 = \mathbf{y}_2 \sim N_p(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}).$$

4.15 Definition: For any positive integer d, χ^2_d is the distribution of $\sum_{i=1}^d Z_i^2$, where $Z_1, \ldots, Z_d$ are independent and identically distributed $N(0, 1)$ random variables.

4.16 Example: Let $Y_1, \ldots, Y_n$ be independent $N(\mu, \sigma^2)$ random variables. Then $\bar{Y}$ and S^2 are independent and $(n-1) \times S^2/\sigma^2 \sim \chi^2_{n-1}$.

In linear model theory, test statistics arise from sums of squares (special cases of quadratic forms) with χ^2 distributions.

4.17 Theorem: If $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma}$ is positive definite, then $(\mathbf{Y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu}) \sim \chi^2_n$.

4.18 Theorem: Let $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ and $\mathbf{P}_{n\times n}$ be symmetric of rank r. Then $Q = (\mathbf{Y} - \boldsymbol{\mu})'\mathbf{P}(\mathbf{Y} - \boldsymbol{\mu})/\sigma^2 \sim \chi^2_r$ if and only if $\mathbf{P}$ is idempotent (i.e. $\mathbf{P}^2 = \mathbf{P}$), and hence a projection.

4.19 Note: Theorem 4.18 says that in the spherically symmetric case $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$, the only quadratic forms with χ^2 distributions are sums of squares, i.e. squared lengths of projections: $\mathbf{x}'\mathbf{P}\mathbf{x} = ||\mathbf{P}\mathbf{x}||^2$.

Theorem 4.22 addresses conditions under which the difference of two χ^2-distributed quadratic forms is χ^2 (to be applied to the ANOVA decomposition of the sum of squares). To prove the theorem, we will need to know when the difference of two projection matrices is a projection matrix.

4.20 Theorem: Assume that $\mathbf{P}_1$ and $\mathbf{P}_2$ are projection matrices and that $\mathbf{P}_1 - \mathbf{P}_2$ is positive semidefinite. Then

(a) $\mathbf{P}_1\mathbf{P}_2 = \mathbf{P}_2\mathbf{P}_1 = \mathbf{P}_2$,

(b) $\mathbf{P}_1 - \mathbf{P}_2$ is a projection matrix.

4.21 Note: The actual interpretation of Theorem 4.20 is:

1. $\mathbf{P}_1$ is a projection onto a linear space Ω.
2. $\mathbf{P}_2$ is a projection onto a subspace ω of Ω.
3. $\mathbf{P}_1 - \mathbf{P}_2$ is a projection onto the orthogonal complement of ω within Ω.

4.22 Theorem: Let $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ and $Q_i = (\mathbf{Y} - \boldsymbol{\mu})'\mathbf{P}_i(\mathbf{Y} - \boldsymbol{\mu})/\sigma^2$, where $\mathbf{P}_i$ is a symmetric $n \times n$ matrix for $i = 1, 2$. If $Q_i \sim \chi^2_{r_i}$ and $Q_1 - Q_2 \geq 0$, then $Q_1 - Q_2$ and Q_2 are independent, and $Q_1 - Q_2 \sim \chi^2_{r_1 - r_2}$.

4.23 Definition: The non-central chi-squared distribution with n degrees of freedom and non-centrality parameter λ, denoted $\chi^2_n(\lambda)$, is defined as the distribution of $\sum_{i=1}^n Z_i^2$, where $Z_1, \ldots, Z_n$ are independent $N(\mu_i, 1)$ random variables, and $\lambda = \sum_{i=1}^n \mu_i^2/2$.

4.24 Note: For any n we have $\chi^2_n(0) \equiv \chi^2_n$, which we refer to as the central chi-square distribution.

4.25 Theorem: If $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \mathbf{I})$, then $\mathbf{Y}'\mathbf{Y}$ has moment generating function

$$M_{\mathbf{Y}'\mathbf{Y}}(t) = (1 - 2t)^{-\frac{n}{2}} \exp\left\{\frac{\boldsymbol{\mu}'\boldsymbol{\mu}}{2}\left[\frac{1}{1 - 2t} - 1\right]\right\}, \quad t < 1/2.$$

4.26 Theorem: Let $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I}_n)$ and $\mathbf{P} = \mathbf{P}'$. Then $\mathbf{P} = \mathbf{P}^2$ of rank r if and only if

$$\mathbf{Y}'\mathbf{P}\mathbf{Y}/\sigma^2 \sim \chi^2_r(\boldsymbol{\mu}'\mathbf{P}\boldsymbol{\mu}/2\sigma^2).$$

4.27 Theorem: If $Y \sim \chi^2(n, \lambda)$, then $E[Y] = n + 2\lambda$, $\mathrm{var}[Y] = 2(n + 4\lambda)$.

4.28 Theorem: If $Y \sim \chi^2_n$ with $n > 2$, then $E[\frac{1}{Y}] = \frac{1}{n-2}$.

4.29 Theorem: $\chi^2_n(\lambda)$, like χ^2_n, has the convolution property.