Ma 3/103
Introduction to Probability and Statistics

KC Border
Winter 2021

Lecture 11: An Introduction to The Multivariate Normal Distribution

Relevant textbook passages:

Larsen–Marx [6]: Section 10.5.

Pitman [7]: Section 6.5.

11.1 The Multivariate Normal distribution

Recall that the Normal $N(\mu, \sigma^2)$ has a density of the form

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

although I don't normally write it in that form.

I am going to use one of two standard definitions of the multivariate normal that makes life simpler. I learned this approach from Dave Grether who told me that he learned it from the late statistician Edwin James George Pitman, who happens to be the father of Jim Pitman, the author of the probability textbook [7] for the course. I have also found this definition is also the one used by Jacod and Protter [5, Definition 16.1, p. 126] and by Rao [8, p. 518].

11.1.1 Definition *Extend the notion of a Normal random variable to include constants as $N(\mu, 0)$ zero-variance (degenerate) random variables.*

*A random vector $\boldsymbol{X} = (X_1, \ldots, X_n) \in \boldsymbol{R}^{\mathrm{n}}$ has a **multivariate Normal distribution** or a **jointly Normal distribution** if for every constant vector $\boldsymbol{w} \in \boldsymbol{R}^{\mathrm{n}}$ the linear combination $\boldsymbol{w}'\boldsymbol{X} = \sum_{i=1}^{n} w_i X_i$ has a univariate Normal distribution.*

Recall Proposition 10.8.1:

10.8.1 Proposition *If X and Y are independent normals with $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\lambda, \tau^2)$, then*

$$X + Y \sim N(\mu + \lambda, \sigma^2 + \tau^2).$$

It has the following Corollary:

11.1.2 Corollary *If $X_1, \ldots, X_n$ are independent Normal random variables, then the random vector*

$$\boldsymbol{X} = (X_1, \ldots, X_n)$$

has a multivariate Normal distribution.

Proof: We need to show that for any constant vector $\boldsymbol{w}$, the linear combination $\boldsymbol{w}'\boldsymbol{X} = \sum_{i=1}^{n} w_i X_i$ has a Normal distribution. But since the X_i's are independent Normals, the $w_i X_i$'s are also independent Normals, so by the Proposition, their sum is a Normal random variable. ∎

So now we know that multivariate Normal random vectors do exist.

11.1.3 Proposition *If $\boldsymbol{X}$ is an n-dimensional multivariate Normal random vector, and $\boldsymbol{A}$ is an $m \times n$ constant matrix, then*

$$\boldsymbol{Y} = \boldsymbol{AX}$$

is an m-dimensional multivariate Normal random vector.

Proof: For a constant $1 \times m$-vector $\boldsymbol{w}$, the linear combination $\boldsymbol{w'Y} = \boldsymbol{w'AX} = (\boldsymbol{Aw})'\boldsymbol{X}$, which is of the form $\boldsymbol{v'X}$ for $\boldsymbol{v} = \boldsymbol{Aw}$, which by hypothesis is univariate Normal. ∎

11.1.4 Proposition *If $\boldsymbol{X} = (X_1, \ldots, X_n)$ has a multivariate Normal distribution, then*

- *Each component X_i has a Normal distribution.*
- *Every subvector of $\boldsymbol{X}$ has a multivariate Normal distribution.*

11.1.5 Definition *Let $\boldsymbol{\mu} = \boldsymbol{E}\,\boldsymbol{X}$ and let*

$$\boldsymbol{\Sigma} = \boldsymbol{E}\big((\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'\big),$$

that is,

$$\sigma_{ij} := \Sigma_{ij} = \boldsymbol{Cov}(X_i, X_j) = \boldsymbol{E}(X_i - \boldsymbol{E}\,X_i)(X_j - \boldsymbol{E}\,X_j)$$

11.1.6 Theorem *If $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then*

$$\boldsymbol{Y} = \boldsymbol{CX} \sim N(\boldsymbol{C\mu}, \boldsymbol{C\Sigma C'}).$$

- Let $\boldsymbol{C}$ be diagonal. Then $\boldsymbol{CX}$ is a linear combination $c_1 X_1 + \cdots + c_n X_n$ of the components and has a (univariate) normal $N(\boldsymbol{C\mu}, \boldsymbol{C\Sigma C'})$ distribution.

- $\boldsymbol{\Sigma}$ is positive semi-definite.

To see this, let $\boldsymbol{w} \in \boldsymbol{R}^{\mathrm{n}}$. Then $\boldsymbol{Var}(\boldsymbol{w'X}) = \boldsymbol{w'\Sigma w}$, but variances are always nonnegative.

The following theorem may be found in Jacod and Protter [5, Theorem 16.1, p. 126–127]. It is used by Anderson [1, § 2.4] as the definition of a multivariate normal.

11.1.7 Proposition *A random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ with nonsingular covariance matrix $\boldsymbol{\Sigma}$ has a multivariate normal distribution if and only if its density is*

$$f(\boldsymbol{x}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sqrt{\det \boldsymbol{\Sigma}}}\, e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}. \tag{2}$$

Note that this agrees with (1) when $n = 1$.

- The multivariate Normal density is constant on ellipsoids of the form

$$(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) = \text{constant}.$$

11.1.8 Corollary *Uncorrelated jointly normal random variables are in fact independent!!*

Proof: If the random variables are uncorrelated, then $\boldsymbol{\Sigma}$ is diagonal. In that case the quadratic form $(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$ reduces to a sum of squares, so the density factors into the product of the marginal densities, which implies independence. ∎

Singular covariance matrices

When a jointly Normal distribution has a singular covariance matrix, then the density does not exist. But if the matrix has rank k, there is a k-dimensional **flat**[1] on which it is possible to define a density of the form given in (2), but we will not make use of that here.

11.2 ⋆ Not jointly normal vectors

If $X_1, \dots, X_n$ are independent Normals, then the vector $(X_1, \dots, X_n)$ is a jointly Normal vector. But if the components are not independent, then the vector may not be jointly Normal. That is, there may be a linear combination of them that does not have a Normal distribution. Here are a couple of examples, the details of which I leave as an exercise.

11.2.1 Example (Example 2, [5, p. 134]) Let Z be a standard Normal and let $a > 0$. Define

$$X = \begin{cases} Z & \text{if } |Z| \leqslant a \\ -Z & \text{if } |Z| > a. \end{cases}$$

Then you can show that X has a standard Normal distribution, $X + Z$ is not constant, but $P(X + Z > 2a) = 0$. Thus the random vecyor (X, Z) is not jointly Normal. □

11.2.2 Example (Exercise 4.47, [4, p. 200]) Let X and Y be independent standard normal random variables. Define the random variable Z by

$$Z = \begin{cases} X & \text{if } XY > 0 \\ -X & \text{if } XY < 0. \end{cases}$$

Then Z is a standard normal random variable, but the random vector (Y, Z) is *not* jointly normal.

I leave the verification as an exercise. (Hint for the second claim: Y and Z always have the same sign.) □

See Casella and Berger [4, Exercises 4.48–4.50, pp. 200–201] for more peculiar examples of pairs of normals that are not jointly normal. Behboodian [3] describes a pair of uncorrelated normal random variables that are not independent—they aren't jointly normal either.

11.3 ⋆ Multivariate Central Limit Theorem

The following vector version of the Central Limit Theorem may be found in Jacod and Protter [5, Theorem 21.3, p. 183].

11.3.1 Multivariate Central Limit Theorem *Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_i \dots$ be independent and identically distributed random k-vectors with common mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Let $\boldsymbol{S}_n = \boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n$. Then*

$$\frac{\boldsymbol{S}_n - \boldsymbol{\mu}}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(\boldsymbol{0}, \boldsymbol{\Sigma}).$$

Note that this result asserts that the limiting distribution is jointly Normal, which is stronger than saying that the distribution of each component is Normal.

One way to prove this result is use the Cramér–Wold device 12.10.5 and the standard Central Limit Theorem 12.5.1. (This statement assumes finite third moments, but the assumption is not necessary. I should update this.)

[1] A **flat** is a translate of a linear subspace. That is, if M is a linear subspace, then $\boldsymbol{x} + M = \{\boldsymbol{x} + \boldsymbol{y} : \boldsymbol{y} \in M\}$ is a flat.

11.4 Multivariate Normal and Chi-square

11.4.1 Proposition *Let $\boldsymbol{X} \sim N(\boldsymbol{0}, \boldsymbol{I}_n)$. Then $\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} \sim \chi^2(k)$ if and only if $\boldsymbol{A}$ is symmetric, idempotent, and has rank k.*

Proof: I'll prove only one half; the half that is most useful in statistics. Assume $\boldsymbol{A}$ is symmetric, idempotent, and has rank k. (Then it is orthogonal projection onto a k-dimensional subspace.) Its eigenvalues are 0 and 1, so it is positive semidefinite. So by the Principal Axis Theorem, there is an orthogonal matrix $\boldsymbol{C}$ such that

$$\boldsymbol{C}'\boldsymbol{A}\boldsymbol{C} = \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

where the λ_i's are the eigenvalues, k of them are 1 and $n-k$ are zero. Setting $\boldsymbol{Y} = \boldsymbol{C}'\boldsymbol{X}$, we see that $\boldsymbol{Y} \sim N(\boldsymbol{0}, \boldsymbol{\Lambda})$. This means that the components of $\boldsymbol{Y}$ are independent. Moreover the Principal Axis Theorem also implies

$$\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} = \boldsymbol{Y}'\boldsymbol{\Lambda}\boldsymbol{Y} = \sum_{i:\lambda_i=1} Y_i^2,$$

which has a $\chi^2(k)$ distribution. ∎

11.4.2 Corollary *Let $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \sigma^2\boldsymbol{I}_n)$. Then*

$$\left(\frac{\boldsymbol{X}-\boldsymbol{\mu}}{\sigma}\right)' \boldsymbol{A} \left(\frac{\boldsymbol{X}-\boldsymbol{\mu}}{\sigma}\right) \sim \chi^2(k)$$

if and only if $\boldsymbol{A}$ is symmetric, idempotent, and has rank k.

Proof: Note that $\frac{\boldsymbol{X}-\boldsymbol{\mu}}{\sigma} \sim N(\boldsymbol{0}, \boldsymbol{I})$. ∎

The following theorem is useful in deriving the distribution of certain test statistics. You can find a proof in C. Radakrishna Rao [8, Item (v), p. 187] or Henri Theil [9, pp. 83–84].

11.4.3 Theorem *Let $\boldsymbol{X} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ and let $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ be symmetric idempotent matrices that satisfy*

$$\boldsymbol{A}_1\boldsymbol{A}_2 = \boldsymbol{A}_2\boldsymbol{A}_1 = \boldsymbol{0}.$$

Then $\boldsymbol{X}'\boldsymbol{A}_1\boldsymbol{X}$ and $\boldsymbol{X}'\boldsymbol{A}_2\boldsymbol{X}$ are independent.

11.5 Independence of Sample Mean and Variance Estimators

11.5.1 Theorem *Let $X_1, \dots, X_n$ be independent and identically distributed Normal $N(\mu, \sigma^2)$ random variables. Let $D_i = X_i - \bar{X}$.*

The sample average $\bar{X}$ and the random vector $(D_1, \dots, D_n)$ are stochastically independent.

N.B. This does not say that each D_i and D_j are independent of each other, rather, they are jointly independent of $\bar{X}$.

Proof: The random vector $\boldsymbol{X} = (X_1, \dots, X_n)$ has a multivariate Normal distribution (Corollary 11.1.2). Therefore the random vector

$$\begin{bmatrix} D_1 \\ \vdots \\ \vdots \\ D_n \\ \bar{X} \end{bmatrix} = \begin{bmatrix} 1-\frac{1}{n} & -\frac{1}{n} & & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & & \ddots & & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & & 1-\frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \cdots & & \frac{1}{n} \end{bmatrix} \begin{bmatrix} X_1 \\ \vdots \\ \vdots \\ X_n \end{bmatrix}$$

is also a multivariate Normal random vector (Theorem 11.1.6).

But by the Covariance Menagerie 9.10.1.15, $\boldsymbol{Cov}(D_i, \bar{X}) = 0$. But for multivariate Normal vectors, this means that $\bar{X}$ and $(D_1, \dots, D_n)$ are stochastically independent. ∎

11.5.2 Corollary *If $X_1, \dots, X_n$ are independent and identically distributed Normal $N(\mu, \sigma^2)$ random variables. Define*

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}.$$

1. $\bar{X} \sim N(\mu, \sigma^2/n)$.
2. *$\bar{X}$ and S^2 are independent.*
3. $\dfrac{(n-1)S^2}{\sigma^2} = \dfrac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n-1)$

Proof:

(1). This is old news.

(2). By Theorem 11.5.1, $\bar{X}$ is independent of $(D_1, \dots, D_n)$. But $S^2 = \sum_i D_i^2/(n-1)$ is a function of $(D_1, \dots, D_n)$, so it too is independent of $\bar{X}$.

(3). Define the standardized version of X_i, Clean this up!!

$$Y_i = \frac{X_i - \mu}{\sigma}, \quad (i = 1, \dots, n), \quad \bar{Y} = \sum_{i=1}^n Y_i/n = (\bar{X} - \mu)/\sigma.$$

Note that for each i,

$$Y_i - \bar{Y} = (X_i - \bar{X})/\sigma,$$

each Y_i is a standard Normal random variable, and the Y_i's are independent. So $\boldsymbol{Y} = (Y_1, \dots, Y_n)$ multivariate Normal with covariance matrix $\boldsymbol{I}$. Let

$$\boldsymbol{v} = \left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}} \right).$$

Note that $\boldsymbol{v}'\boldsymbol{v} = 1$. Now create an orthogonal matrix $\boldsymbol{B}$ that has $\boldsymbol{v}$ as its last row. We can always do this. (Think of the Gram–Schmidt procedure.)

Define the transformed variables

$$\boldsymbol{Z} = \boldsymbol{B}\boldsymbol{Y}.$$

Since $\boldsymbol{Y}$ is multivariate Normal, therefore so is $\boldsymbol{Z}$. By Proposition 11.1.3, the covariance matrices satisfy

$$\boldsymbol{Var}\,\boldsymbol{Z} = \boldsymbol{B}(\boldsymbol{Var}\,\boldsymbol{Y})\boldsymbol{B}' = \boldsymbol{B}\boldsymbol{B}' = \boldsymbol{B}\boldsymbol{B}^{-1} = \boldsymbol{I},$$

so $\boldsymbol{Z}$ is a vector of independent standard Normal random variables.

By Proposition S2.5.2, multiplication by $\boldsymbol{B}$ preserves norms, so

$$\sum_{i=1}^n Z_i^2 = \sum_{i=1}^n Y_i^2. \tag{3}$$

But Z_n is the dot product of the last row of $\boldsymbol{B}$ with $\boldsymbol{Y}$, or

$$Z_n = \boldsymbol{v} \cdot \boldsymbol{Y} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i = \sqrt{n}\bar{Y}. \tag{4}$$

So combining (3) and (4), we have

$$\sum_{i=1}^n Y_i^2 = \sum_{i=1}^{n-1} Z_i^2 + n\bar{Y}^2. \tag{5}$$

On the other hand, by the Pythagorean Theorem for Data S2.4.4, we can write

$$\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 + n\bar{Y}^2, \tag{6}$$

Combining (5) and (6) implies

$$\sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n-1} Z_i^2. \tag{7}$$

Now rewrite this in terms of the X_i's:

$$\sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X})^2. \tag{8}$$

Combining (7) and (8) shows that

$$\frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n-1} Z_i^2$$

where Z_i are independent standard Normals. In other words, $\frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X})^2$ has a $\chi^2(n-1)$ distribution. ∎

Bibliography

[1] T. W. Anderson. 1958. *An introduction to multivariate statistical analysis*. A Wiley Publication in Mathematical Statistics. New York: Wiley.

[2] S. Axler. 1997. *Linear algebra done right*, 2d. ed. Undergraduate Texts in Mathematics. New York: Springer.

[3] J. Behboodian. 1990. Examples of uncorrelated dependent random variables using a bivariate mixture. *The American Statistician* 44(3):218. http://www.jstor.org/stable/2685339

[4] G. Casella and R. L. Berger. 2002. *Statistical inference*, 2d. ed. Belmont, California: Brooks/Cole Cengage Learning.

[5] J. Jacod and P. Protter. 2004. *Probability essentials*, 2d. ed. Berlin and Heidelberg: Springer.

[6] R. J. Larsen and M. L. Marx. 2012. *An introduction to mathematical statistics and its applications*, fifth ed. Boston: Prentice Hall.

[7] J. Pitman. 1993. *Probability*. Springer Texts in Statistics. New York, Berlin, and Heidelberg: Springer.

[8] C. R. Rao. 1973. *Linear statistical inference and its applications*, 2d. ed. Wiley Series in Probability and Mathematical Statistics. New York: Wiley.

[9] H. Theil. 1971. *Principles of econometrics*. New York: Wiley.