Multivariate Normal Distribution

Olivia R. McClammy

For this presentation, I made references to "Introduction to Probability" by Joseph K. Blitzstein and Jessica Hwang. I would also like to thank Professor Everson for his instruction on this problem.

Multivariate Normal Distribution Definition

4. **Multivariate Normal Variables** (Blitzstein 7.5)
   Definition: A random vector $(X_1, \ldots, X_n)$ follows a *Multivariate Normal* (MVN) distribution if $t_1X_1 + \ldots + t_nX_n$ follows a Normal distribution for any choice of constants $t_1, \ldots, t_n$.

4. **Multivariate Normal Variables** (Blitzstein 7.5)

Definition: A random vector $(X_1, \ldots, X_n)$ follows a *Multivariate Normal* (MVN) distribution if $t_1X_1 + \ldots + t_nX_n$ follows a Normal distribution for any choice of constants $t_1, \ldots, t_n$.

a) Explain why this implies that any sample of n *independent* Normal random variables follows ~~an~~ n dimensional MVN distribution.

Goal: $\{X_1, X_2, \ldots, X_n\}$ is a MVN distribution if and only if $\sum_{i=1}^{n} t_i X_i$ must be Normal

Given $Z_1, Z_2, \ldots, Z_n$ iid $N(0,1)$

We know from the MGF's proof that:

A sum of independent Normal variables is Normal

at $c_1 Z_1 + c_2 Z_2 + c_3 Z_3 + \ldots + c_n Z_n \sim$ Normal

Given any vector $X_1, \ldots, X_m$

Where X_j is a linear combination of Z_i's $j = 1, \ldots, m$

(any vector of m independent Normal variables $X_1, \ldots, X_m$ is a linear transformation of $Z_1, \ldots, Z_n \overset{iid}{\sim} N(0,1)$)

is an MVN vector because $t_1 X_1 + t_2 X_2 + \ldots + t_m X_m$

is a linear combination of $Z_1, Z_2, \ldots, Z_n$

Under the condition that $X_1, \ldots, X_n$ has a proper pdf if and only if $m \le n$ ← conditional

MGF Proof (PSU)

b) Explain why, if X_i is part of a MVN vector, then X_i is Normal. Also give an example of variables X and Y that have marginal Normal distributions, but that are not part of a multivariate Normal vector.

Goal: If X_i part of a MVN vector, then X_i is Normal

X_i is part of a MVN vector: $t_1 X_1 + t_2 X_2 + \ldots + t_i X_i + \ldots + t_n x_n$

We know that $t_1, \ldots, t_n$ is any choice of constants (given by MVN distribution definition)

Choose $t_i = 1$, t_j's $= 0$

Then, $0(x_1) + 0(x_2) + \ldots + 1(x_i) + \ldots + 0(x_n) = \boxed{X_i}$

Therefore X_i is Normal (becomes a univariate Normal)

References Blitzstein textbook (Page 310)

b) Explain why, if X_i is part of a MVN vector, then X_i is Normal. Also give an example of variables X and Y that have marginal Normal distributions, but that are not part of a multivariate Normal vector.

Part 2:

Goal: Find an example of two variables (X and Y) that have Normal Distributions but $X+Y$ is not multivariate Normal Vector

$X \sim N(0,1)$

$$S = \begin{cases} 1 & \mathbb{P}(S=1) = \tfrac{1}{2} \\ -1 & \mathbb{P}(S=-1) = \tfrac{1}{2} \end{cases}$$

$Y = SX$ $\sim$ Normal (why S has to be $-1, 1$)

Is $X+Y$ a multivariate Normal vector?

For $X+Y$ to be a multivariate normal vector, $X+Y$ would have to follow a normal distribution $\leftarrow$ definition of MVN distribution

2 cases: $S=1$, $X+Y=2X$ $\leftarrow$ This is continuous since $X \sim N(0,1)$

$S=-1$, $\boxed{X+Y=0}$ $\leftarrow$ $\mathbb{P}(X+Y=0) = \tfrac{1}{2}$ which is discrete

Since the probabilities are partly continuous and partly discrete, cannot be Normal, therefore, $X+Y$ are not a multivariate Normal vector

References Blitzstein textbook (Page 310)

c) The n-dimensional multivariate Normal density with $n \times 1$ mean vector $\boldsymbol{\mu}$ and $n \times n$ covariance matrix $\mathbf{V}$ has a proper joint pdf if and only if $\mathbf{V}$ is positive definite, and hence invertible. We write $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \mathbf{V})$, and the joint pdf is

$$f(\mathbf{x}) = (2\pi)^{-n/2}|\mathbf{V}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

Show that this is a generalization of the univariate Normal density ($n = 1$) and that the vector of $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$ has a joint pdf of this form (what are $\boldsymbol{\mu}$ and $\mathbf{V}$?).

Linear Algebra Refresher:

Transpose:

$$X = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \quad X^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

2×3 $\qquad$ 3×2

Determinant:

$$|X| = \begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = \cancel{\frac{1}{(1)(4)-(2)(3)}}$$

$= (1)(4) - (2)(3)$

$= ad - bc$

Identity Matrix (I):

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

c) The n-dimensional multivariate Normal density with $n \times 1$ mean vector $\boldsymbol{\mu}$ and $n \times n$ covariance matrix $\mathbf{V}$ has a proper joint pdf if and only if $\mathbf{V}$ is positive definite, and hence invertible. We write $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \mathbf{V})$, and the joint pdf is

$$f(\mathbf{x}) = (2\pi)^{-n/2}|\mathbf{V}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

Show that this is a generalization of the univariate Normal density ($n = 1$) and that the vector of $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$ has a joint pdf of this form (what are $\boldsymbol{\mu}$ and $\mathbf{V}$?).

Goal: Starting with univariate Normal Distribution: $\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{(x-u)}{\sigma}\right)^2}$

to $(2\pi)^{-n/2} |V|^{-1/2} \exp\left[-\frac{1}{2}(x-u)^T V^{-1}(x-u)\right]$ and find u and V.

Start: $\frac{1}{\sqrt{2\pi}\,\sigma} e^{-1/2\left(\frac{(x-u)}{\sigma}\right)^2}$

We know: The multivariate Normal density is the product of n Univariate Normal densities

$(x_1, \ldots, x_n) \overset{iid}{\sim} N(u, \sigma^2)$

$$f_{x_1,\ldots,x_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x_i-u)^2}{\sigma^2}} \right)$$

$$= (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp\left[-1/2 \sum (x_i-u)^2/\sigma^2\right]$$

$|V|^{-1/2}$: $|V| = \begin{vmatrix} \sigma_1^2 & & 0 \\ 0 & \ddots & \sigma_n^2 \end{vmatrix}$ from $(\sigma^2)^{-n/2}$ since covariance matrix is $\sigma^2 I$ and $\det = (\sigma^2)^n$

$(x-u)^T = \sum (x_i - u)^2 / \sigma^2 \leftarrow V^{-1}$

Helpful Facts:

If $\underset{1\times 1}{X = [x]}$, then $\underset{1\times 1}{X^T = [x]}$ and $(x-u)^2 = (x-u)^T(x-u)$

$$\frac{\sum (x_i - u)^2}{\sigma} = \underbrace{(x_1 - u_1, \dots, x_n - u_n)}_{(x-u)^T} \underbrace{\begin{pmatrix} 1/\sigma^2 & & 0 \\ & \ddots & \\ 0 & & 1/\sigma^2 \end{pmatrix}}_{V^{-1}} \underbrace{\begin{pmatrix} x_i - u \\ \vdots \\ x_i - u \end{pmatrix}}_{(x-u)}$$

$$= (x_i - u)^T V^{-1} (x-u)$$

Then we get: $(2\pi)^{-n/2} |V|^{-1/2} \exp\left[-1/2 (x-u)^T V^{-1} (x-u)\right]$

From the problem we know:

We have a $n \times 1$ mean vector $u = u \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

We have a $n \times n$ covariance matrix $\underset{n\times n}{V} = \sigma^2 \underset{n\times n}{I} = \sigma^2 \begin{pmatrix} 1_1 & & 0 \\ & \ddots & \\ 0 & & 1_n \end{pmatrix}$

d) Let $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ represent a multivariate Normal vector of dimension $m+n$, and suppose $\mathrm{Cov}(X_i, Y_j) = 0$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Show that the joint density of the X's and Y's factors into a joint pdf for X_i's and a joint pdf for the Y_j's, meaning these are independent random vectors. State the general result about correlation and independence for elements of a multivariate Normal vector.

Goal: We want to show that $f(x, y) = f(x) f(y)$ to mean that there are independent, random vectors while stating the general result of correlation and independence.

Start: $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ is a multivariate normal vector; dimension $m+n$

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \\ Y_1 \\ \vdots \\ Y_m \end{pmatrix} \sim N\left(\begin{pmatrix} \begin{pmatrix} u_{X_1} \\ \vdots \\ u_{X_m} \end{pmatrix} \\ \begin{pmatrix} u_{Y_1} \\ \vdots \\ u_{Y_m} \end{pmatrix} \end{pmatrix}, \begin{pmatrix} \underset{n \times n}{V_X} & \underset{n \times m}{0} \\ \underset{m \times n}{0} & \underset{m \times m}{V_Y} \end{pmatrix} \right)$$

X

$\mathrm{Cov} = 0$

$\mathrm{Cov} = 0$

Y

There are 0's because we know $\mathrm{Cov}(X_i, Y_j) = 0$ (from the problem)

$\mathrm{Cov}(X_i, Y_j) = 0$ $\quad i = 1, \ldots, n$, $j = 1, \ldots, m$, We can assume $V_X, V_Y \longleftrightarrow V$ are invertible ($\det \neq 0$)

d) Let $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ represent a multivariate Normal vector of dimension $m+n$, and suppose $\mathrm{Cov}(X_i, Y_j) = 0$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Show that the joint density of the X's and Y's factors into a joint pdf for X_i's and a joint pdf for the Y_j's, meaning these are independent random vectors. State the general result about correlation and independence for elements of a multivariate Normal vector.

$$V = \begin{array}{c|c} V_1 & 0 \\ \hline 0 & V_2 \end{array}$$

$$|V| = |V_1||V_2|$$

$$V^{-1} = \begin{pmatrix} V_1^{-1} & 0 \\ 0 & V_2^{-1} \end{pmatrix}$$

$$f_{XY} = (2\pi)^{\frac{-(n+m)}{2}} |V|^{-1/2} \exp\left[-\tfrac{1}{2}\left((x-u_x)^T (y-u_y)^T\right) V^{-1} \begin{pmatrix} x-u_x \\ \vdots \\ y-u_y \end{pmatrix}\right.$$ ← joint density X and Y

$$\downarrow \rho = 0$$

$$f(x)f(y) = \underbrace{(2\pi)^{-n/2}|V_1|^{-1/2}}_{X}\underbrace{(2\pi)^{-m/2}|V_2|^{-1/2}}_{Y}\exp\left[\underbrace{-\tfrac{1}{2}(x-u_x)^T V_1^{-1}(x-u_x)}_{x} + \underbrace{-\tfrac{1}{2}(y-u_y)^T V_2^{-1}(y-u_y)}_{Y}\right.$$

$f_{XY}(x,y) = f(x) f_Y(y)$ if and only if X, Y are independent and Correlation $= 0$

Example

Example: $n = m = 1$

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2\left(\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix} \right)$$

$$f_{(X,Y)}(x,y) = \frac{1}{\sqrt{2\pi}\, V_1^{1/2}} e^{-\frac{(x-u_x)^2}{2V_1}} \frac{1}{\sqrt{2\pi}\, V_2^{1/2}} e^{-\frac{(y-u_y)^2}{2V_2}}$$

$$\begin{vmatrix} V_1 & 0 \\ 0 & V_2 \end{vmatrix} = V_1 V_2 \qquad \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix}^{-1} = \begin{bmatrix} V_1^{-1} & 0 \\ 0 & V_2^{-1} \end{bmatrix}$$

Application of Multivariate Normal Distributions: Stock Return

Financial Economics: Stock market and how to estimate stock returns

"A Test for Multivariate Normality in Stock Returns" by Richardson and Smith in 1992 aimed to see if stock returns could be a multivariate normal distribution (previously stock return was not able to be described as univariate normal distributions)

Discussed dependence, cross moments, multivariate normal, multivariate time series, and generalized method of moments estimator to run tests on stock/asset returns and market model residuals.

Determined "non normality in both the marginal and joint distributions of these variables." (Richardson and Smith, 1992)

Richardson and Smith Paper