Fall 2019 Statistics 201A (Introduction to Probability at an advanced level) - Lecture Twenty Four

Aditya Guntuboyina

03 December, 2019

1 Moment Generating Functions of Random Vectors

We shall next move to the last topic of the class: the multivariate normal distribution. For this, it is helpful to know about moment generating functions of random vectors.

The Moment Generating Function of an $n \times 1$ random vector Y is defined as

$$M_Y(a) := \mathbb{E}e^{a^T Y}$$

for every $a \in \mathbb{R}^n$ for which the expectation exists. Note that when $a = (0, \dots, 0)^T$ is the zero vector, it is easy to see that $M_Y(a) = 1$.

Just like in the univariate case, Moment Generating Functions determine distributions when they exist in a neighbourhood of $a = 0$.

Moment Generating Functions behave very nicely in the presence of independence. Suppose $Y_{(1)}$ and $Y_{(2)}$ are two random vectors and let $Y = (Y_{(1)}^T, Y_{(2)}^T)^T$ be the vector obtained by putting $Y_{(1)}$ and $Y_{(2)}$ together in a single column vector. Then $Y_{(1)}$ **and** $Y_{(2)}$ **are independent if and only if**

$$M_Y(a) = M_{Y_{(1)}}(a_{(1)})M_{Y_{(2)}}(a_{(2)}) \qquad \text{for every } a = (a_{(1)}^T, a_{(2)}^T)^T \in \mathbb{R}^n.$$

Thus under independence, the MGF factorizes and conversely, when the MGF factorizes, we have independence.

2 The Multivariate Normal Distribution

The multivariate normal distribution is defined in the following way.

Definition 2.1. *A random vector $Y = (Y_1, \dots, Y_n)^T$ is said to have the multivariate normal distribution if every linear function $a^T Y$ of Y has the univariate normal distribution.*

Remark 2.1. *It is important to emphasize that for $Y = (Y_1, \dots, Y_n)^T$ to be multivariate normal,* ***every*** *linear function $a^T Y = a_1 Y_1 + \dots a_n Y_n$ needs to be univariate normal. It is not enough for example to just have each Y_i to be univariate normal. It is very easy to construct examples where each Y_i is univariate normal but $a_1 Y_1 + \dots + a_n Y_n$ is not univariate normal for many vectors $(a_1, \dots, a_n)^T$. For example, suppose that $Y_1 \sim N(0, 1)$ and that $Y_2 = \xi Y_1$ where ξ is a discrete random variable taking the two values 1 and -1 with probability $1/2$ and ξ is independent of Y_1. Then it is easy to see that*

$$Y_2 | \xi = 1 \sim N(0, 1) \quad \textit{and} \quad Y_2 | \xi = -1 \sim N(0, 1).$$

This means therefore that $Y_2 \sim N(0,1)$ (and that Y_2 is independent of ξ). Note however that $Y_1 + Y_2$ is not normal as

$$\mathbb{P}\{Y_1 + Y_2 = 0\} = \mathbb{P}\{\xi = 1\} = \frac{1}{2}.$$

Thus, in this example, even though Y_1 and Y_2 are both $N(0,1)$, the vector $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$ is not multivariate normal.

Example 2.2. *We have seen earlier in the class that if $Z_1, \dots, Z_n$ are **independent** and univariate normal, then $a_1 Z_1 + \dots a_n Z_n$ is normal for every $a_1, \dots, a_n$. Therefore a random vector $Z = (Z_1, \dots, Z_n)^T$ that is made up of **independent** Normal random variables has the multivariate normal distribution. In fact, we shall show below that if Y has a multivariate normal distribution, then it should necessarily be the case that Y is a linear function of a random vector Z that is made of independent univariate normal random variables.*

2.1 Moment Generating Function of a Multivariate Normal

Suppose $Y = (Y_1, \dots, Y_n)^T$ is multivariate normal. Let $\mu = \mathbb{E}(Y)$ and $\Sigma = Cov(Y)$ be the mean vector and covariance matrix of Y respectively. Then, as a direct consequence of the definition of multivariate normality, it follows that the MGF of Y equals

$$M_Y(a) = \mathbb{E}(e^{a^T Y}) = \exp\left(a^T \mu + \frac{1}{2} a^T \Sigma a\right). \tag{1}$$

To see why this is true, note that by definition of multivariate normality, $a^T Y$ is univariate normal. Now the mean and variance of $a^T Y$ are given by

$$\mathbb{E}(a^T Y) = a^T \mu \quad \text{and} \quad Var(a^T Y) = a^T Cov(Y) a = a^T \Sigma a$$

so that

$$a^T Y \sim N(a^T \mu, a^T \Sigma a) \qquad \text{for every } a \in \mathbb{R}^n.$$

Then (1) directly follows from the formula for the MGF of a univariate normal.

Note that the MGF of Y (given by (1)) only depends on the mean vector μ and the covariance matrix Σ of Y. Thus the distribution of every multivariate normal vector Y is characterized by the mean vector μ and covariance Σ. We therefore use the notation $N_n(\mu, \Sigma)$ for the multivariate normal distribution with mean μ and covariance Σ.

2.2 Connection to i.i.d $N(0,1)$ random variables

Suppose that the covariance matrix Σ of Y is positive definite. Let A be an invertible $n \times n$ matrix so that $\Sigma = AA^T$ (such matrices exist; for example, one can take A to be the unique positive definite square root of Σ; see e.g., `https://yutsumura.com/a-positive-definite-matrix-has-a-unique-positive-definite-square-root/`).

Let $Z := A^{-1}(Y - \mu)$. The formula (1) allows the computation of the MGF of Z as follows:

$$\begin{aligned}
M_Z(a) &= \mathbb{E}e^{a^T Z} \\
&= \mathbb{E}\exp\left(a^T A^{-1}(Y-\mu)\right) \\
&= \exp(-a^T A^{-1}\mu)\mathbb{E}\exp\left(a^T A^{-1} Y\right) \\
&= \exp(-a^T A^{-1}\mu) M_Y((A^{-1})^T a) \\
&= \exp(-a^T A^{-1}\mu)\exp\left(a^T A^{-1}\mu + \frac{1}{2}(a^T A^{-1})\Sigma((A^{-1})^T a)\right) \\
&= \exp(-a^T A^{-1}\mu)\exp\left(a^T A^{-1}\mu + \frac{1}{2}(a^T A^{-1})AA^T((A^{-1})^T a)\right) \\
&= \exp\left(\frac{1}{2}a^T a\right) = \prod_{i=1}^{n}\exp(a_i^2/2).
\end{aligned}$$

The right hand side above is clearly the MGF of a random vector having n i.i.d standard normal random variables. Thus because MGFs uniquely determine distributions, we conclude that $Z = (Z_1, \dots, Z_n)^T$ has independent standard normal random variables. We have thus proved that: **If $Y \sim N_n(\mu, \Sigma)$ and Σ is p.d, then the components $Z_1, \dots Z_n$ of $Z = A^{-1}(Y - \mu)$ are independent standard normal random variables**. This implies that Y **equals the linear combination $Y = AZ + \mu$ of a random vector Z consisting of independent $N(0,1)$ random variables**. Here A is any invertible matrix for which $AA^T = \Sigma$.

3 Joint Density of the Multivariate Normal Distribution

Suppose $Y = (Y_1, \dots, Y_n)^T$ is a random vector that has the multivariate normal distribution. What then is the joint density of $Y_1, \dots, Y_n$?

Let $\mu = \mathbb{E}(Y)$ and $\Sigma = Cov(Y)$ be the mean vector and covariance matrix of Y respectively. For Y to have a joint density, we need to assume that Σ is positive definite. We have then seen in the previous section that the components $Z_1, \dots, Z_n$ of Z are independent standard normal random variables where

$$Z = A^{-1}(Y - \mu) \text{ and } AA^T = \Sigma.$$

Because $Z_1, \dots, Z_n$ are independent standard normals, their joint density equals

$$f_{Z_1,\dots,Z_n}(z_1, \dots, z_n) = (2\pi)^{-n/2}\prod_{i=1}^{n} e^{-z_i^2/2} = (2\pi)^{-n/2}\exp\left(-\frac{1}{2}z^T z\right)$$

where $z = (z_1, \dots, z_n)^T$.

Using the above formula and the fact that $Y = \mu + AZ$, we can compute the joint density of $Y_1, \dots, Y_n$ via the Jacobian formula. This gives

$$\begin{aligned}
f_{Y_1,\dots,Y_n}(y_1, \dots, y_n) &= f_{Z_1,\dots,Z_n}(A^{-1}(y-\mu))\det(A^{-1}) \\
&= \frac{1}{(2\pi)^{n/2}\sqrt{\det(\Sigma)}}\exp\left(\frac{-1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu)\right)
\end{aligned}$$

where $y = (y_1, \dots, y_n)^T$.

4 Properties of Multivariate Normal Random Variables

Suppose $Y = (Y_1, \dots, Y_n)^T \sim N_n(\mu, \Sigma)$. Note then that μ is the mean vector $\mathbb{E}(Y)$ of Y and Σ is the covariance matrix $Cov(Y)$. The following properties are very important.

1. **Linear Functions of Y are also multivariate normal**: If A is an $m \times n$ deterministic matrix and c is an $m \times 1$ deterministic vector, then $AY + c \sim N_m(A\mu + c, A\Sigma A^T)$.

   Reason: Every linear function of $AY + c$ is obviously also a linear function of Y and, thus, this fact follows from the definition of the multivariate normal distribution.

2. If Y is multivariate normal, then every random vector formed by taking a subset of the components of Y is also multivariate normal.

   Reason: Follows from the previous fact.

3. **Independence is the same as Uncorrelatedness**: If $Y_{(1)}$ and $Y_{(2)}$ are two random vectors such that $Y = (Y_{(1)}^T, Y_{(2)}^T)^T$ is multivariate normal. Then $Y_{(1)}$ and $Y_{(2)}$ are independent if and only if $Cov(Y_{(1)}, Y_{(2)}) = 0$.

   Reason: The fact that independence implies $Cov(Y_{(1)}, Y_{(2)}) = 0$ is obvious and does not require any normality. The key is the other implication that zero covariance implies independence. For this, it is enough to show that the MGF of Y equals the product of the MGFs of $Y_{(1)}$ and $Y_{(2)}$. The MGF of Y equals
$$M_Y(a) = \exp\left(a^T\mu + \frac{1}{2}a^T\Sigma a\right)$$
   where $\Sigma = Cov(Y)$.

   Note that $Y_{(1)}$ and $Y_{(2)}$ are also multivariate normal so that
$$M_{Y_{(i)}}(a_{(i)}) = \exp\left(a_{(i)}^T\mu_{(i)} + \frac{1}{2}a_{(i)}^T\Sigma_{ii}a_{(i)}\right) \qquad \text{for } i = 1, 2$$
   where
$$\mu_{(i)} := \mathbb{E}(Y_{(i)}) \quad \text{and} \quad \Sigma_{ii} := Cov(Y_{(i)}).$$
   Now if $\Sigma_{12} := Cov(Y_{(1)}, Y_{(2)})$ and $\Sigma_{21} = Cov(Y_{(2)}, Y_{(1)}) = \Sigma_{12}^T$, then observe that
$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{pmatrix}$$
   As a result, if $a = (a_{(1)}, a_{(2)})^T$, then
$$a^T\Sigma a = a_{(1)}^T\Sigma_{11}a_{(1)} + a_{(2)}^T\Sigma_{22}a_{(2)} + 2a_{(1)}^T\Sigma_{12}a_{(2)}.$$
   Under the assumption that $\Sigma_{12} = 0$, we can therefore write
$$a^T\Sigma a = a_{(1)}^T\Sigma_{11}a_{(1)} + a_{(2)}^T\Sigma_{22}a_{(2)}$$
   from which it follows that
$$M_Y(a) = M_{Y_{(1)}}(a_{(1)})M_{Y_{(2)}}(a_{(2)}).$$
   Because the MGF of $Y = (Y_{(1)}, Y_{(2)})^T$ factorizes into the product of the MGF of $Y_{(1)}$ and the MGF of $Y_{(2)}$, it follows that $Y_{(1)}$ and $Y_{(2)}$ are independent. Thus under the assumption of multivariate normality of $(Y_{(1)}, Y_{(2)})^T$, uncorrelatedness is the same as independence.

4. Suppose $Y = (Y_1, \ldots, Y_n)^T$ is a multivariate normal random vector. Then two components Y_i and Y_j are independent if and only if $\Sigma_{ij} = 0$ where $\Sigma = Cov(Y)$.

   Reason: Follows directly from the previous three facts.

5. **Independence of linear functions can be checked by multiplying matrices**: Suppose Y is multivariate normal. Then AY and BY are independent if and only if $A\Sigma B^T = 0$.

   Reason: Note first that
$$\begin{pmatrix} AY \\ BY \end{pmatrix} = \begin{pmatrix} A \\ B \end{pmatrix} Y$$
   is multivariate normal. Therefore AY and BY are independent if and only if $Cov(AY, BY) = 0$. The claimed assertion then follows from the observation that $Cov(AY, BY) = A\Sigma B^T$.