5. Multivariate Normal Distribution

1. Introduction

An n-dimensional random vector X is said to have (multivariate) normal distribution with parameters μ and Σ, and denoted by $X \sim \mathbf{N}(\mu, \Sigma)$ (or $\mathbf{N}_n(\mu, \Sigma)$ to emphasize the dimension of X), if it has probability density

$$p(x) = \frac{1}{(2\pi)^{n/2}} (\det \Sigma)^{-1/2} \exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right)$$

where $\mu \in \mathbf{R}^n$ and Σ is an $n \times n$ positive definite matrix.

Let Z be an n-vector of independent standard normal variates. Clearly, $Z \sim \mathbf{N}_n(0, I)$. The following lemma shows that any normal random vector can be written as a linear transformation of Z.

Lemma 1 *Let Z be defined as above and*

$$X = \mu + \Sigma^{1/2} Z$$

Then

$$X \sim \mathbf{N}(\mu, \Sigma)$$

Proof Notice that the density of Z is given by

$$p(z) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} z'z\right)$$

and the Jocobian of the transformation $z \mapsto x = \mu + \Sigma^{1/2} z$ is

$$(\det \Sigma)^{-1/2}$$

The stated result then follows immediately. ■

The representation of a normal vector as an affine function of a random vector consisting of independent standard normals is very useful. It is indeed straightforward from such representation that

Corollary 2 *Let* $X \sim \mathbf{N}(\mu, \Sigma)$. *Then*

$$\mathbf{E}\,(X) = \mu \quad \text{and} \quad \text{var}\,(X) = \Sigma$$

Proof Due to Lemma 1, we can write $X = \mu + \Sigma^{1/2} Z$, where Z is a random vector of independent standard normals. We have

$$\begin{aligned} \mathbf{E}\,(X) &= \mu + \Sigma^{1/2}\mathbf{E}(Z) \;=\; \mu \\ \text{var}\,(X) &= \Sigma^{1/2}\text{var}(Z)\,\Sigma^{1/2} \;=\; \Sigma \end{aligned}$$

as was to be shown. ■

We may also easily derive the characteristic function of multivariate normal distribution, as shown below.

Corollary 3 *Let* $X \sim \mathbf{N}(\mu, \Sigma)$. *Then the characteristic function* φ_x *of* X *is given by*

$$\varphi_x(t) = \exp\left(i\mu' t - \frac{1}{2} t' \Sigma t\right)$$

Proof Let Z be defined as above, and denote by φ_z the characteristic function of Z. Also, let

$$t = (t_1, \ldots, t_n)' \quad \text{and} \quad Z = (Z_1, \ldots, Z_n)'$$

From the independence of $Z_1, \ldots, Z_n$, we have

$$\begin{aligned} \varphi_z(t) &= \mathbf{E}(e^{it'Z}) \\ &= \mathbf{E}(e^{it_1 Z_1}) \cdots \mathbf{E}(e^{it_n Z_n}) \\ &= \exp\left(-\frac{1}{2} t' t\right) \end{aligned}$$

It follows that

$$\begin{aligned} \varphi_x(t) &= \mathbf{E}(e^{it'X}) \\ &= e^{it'\mu}\varphi_z(\Sigma^{1/2} t) \\ &= \exp\left(it'\mu - \frac{1}{2} t' \Sigma t\right) \end{aligned}$$

as is required to be shown. ■

Remarks

(a) Any linear combination of the components in a multivariate normal variate is normal, since it can be written as a linear combination of independent normals. This is the property that distinguishes multivariate normal vectors from vectors of univariate normals.

(b) We have thus far assumed that Σ is nonsingular. If Σ is singular, then there exists some $c \in \mathbf{R}^n$ such that

$$\text{var}\,(c'X) = c'\Sigma\, c = 0$$

This implies that the distribution of X is concentrated on a subset of $\mathbf{R}^n$. We often say that the distribution of X is degenerate in this case.

2. Marginal and Conditional Distributions

Throughout this section, we let $X \sim \mathbf{N}(\mu, \Sigma)$. Partition X as

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

where X_1 and X_2 are, respectively, n_1- and n_2-dimensional. Let mean and variance of X be partitioned conformably as

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

The following lemma shows that any affine transformation of a multivariate normal variate is normal.

Lemma 4 *If $Y = AX + b$, then $Y \sim \mathbf{N}(A\mu + b, A\Sigma A')$.*

Proof The characteristic function of Y is given by

$$\begin{aligned} \varphi_y(t) &= \mathbf{E}(e^{it'Y}) \\ &= e^{it'b}\mathbf{E}(e^{it'AX}) \\ &= e^{it'b}\varphi_x(A't) \\ &= e^{it'b}\exp\left(it'A\mu - \frac{1}{2}t'A\Sigma A't\right) \\ &= \exp\left(it'(A\mu + b) - \frac{1}{2}t'A\Sigma A't\right) \end{aligned}$$

which shows that the distribution of Y is multivariate normal with given mean and variance. ∎

We have as a special case that

Corollary 5

$$X_1 \sim \mathbf{N}(\mu_1, \Sigma_{11})$$

Proof Apply Lemma 4 with $A = (I_{n_1}, 0)$ and $b = 0$. ∎

Corollary 5 shows that the marginal distribution of multivariate normal distribution is also normal.

Theorem 6 *X_1 and X_2 are independent if and only if $\Sigma_{12} = 0$.*

Proof The *only if* part is obvious. To prove *if* part, let $\Sigma_{12} = 0$ and write

$$\Sigma = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}$$

Denote by $p(x_1)$ and $p(x_2)$ the probability densities of X_1 and X_2. It follows that

$$\begin{aligned} p(x) &= (2\pi)^{-n/2}(\det\Sigma)^{-1/2}\exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right) \\ &= (2\pi)^{-n_1/2}(\det\Sigma_{11})^{-1/2}\exp\left(-\frac{1}{2}(x_1-\mu_1)'\Sigma_{11}^{-1}(x_1-\mu_1)\right) \\ &\quad \cdot (2\pi)^{-n_2/2}(\det\Sigma_{22})^{-1/2}\exp\left(-\frac{1}{2}(x_2-\mu_2)'\Sigma_{22}^{-1}(x_2-\mu_2)\right) \\ &= p(x_1)p(x_2) \end{aligned}$$

and therefore X_1 and X_2 are independent. ∎

Theorem 7 *The conditional distribution of X_1 given X_2 is*

$$\mathbf{N}(\mu_{1\cdot2}, \Sigma_{11\cdot2})$$

where

$$\begin{aligned} \mu_{1\cdot2} &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2-\mu_2) \\ \Sigma_{11\cdot2} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \end{aligned}$$

Proof Consider a random vector given by

$$\begin{pmatrix} X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2 \\ X_2 \end{pmatrix} = \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

which is normal as a linear transformation of a normal random vector X. The two sub-vectors $X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2$ and X_2 are uncorrelated, and therefore independent.

Write

$$X_1 = \left(X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2\right) + \Sigma_{12}\Sigma_{22}^{-1}X_2$$

The first term is independent of X_2. Its conditional distribution given X_2 is consequently the same as its unconditional distribution, which is normal with mean and variance

$$\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2 \quad \text{and} \quad \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

The second term can be treated as constant when X_2 is given. It therefore just shifts the mean of conditional distribution of X_1 given X_2. ■

Remarks

(a) The conditional mean given X_2 is linear in X_2.

(b) The conditional variance given X_2 does not depend on X_2.

3. Quadratic Forms

We consider the distribution of the quadratic form $X'AX$ in normal random vector X, defined with nonrandom matrix A. It is well known from elementary statistics that for a vector $Z = (Z_1, \ldots, Z_n)'$ of independent standard normals

$$Z'Z = \sum_{i=1}^{n} Z_i^2 \sim \chi_n^2$$

This is generalized below.

Proposition 8 *Let* $X \sim \mathbf{N}_n(0, \Sigma)$. *Then*

$$X'\Sigma^{-1}X \sim \chi_n^2$$

Proof Notice that $\Sigma^{-1/2}X \sim \mathbf{N}(0, I)$. ■

Moreover, we have

Theorem 9 *Let $Z \sim \mathbf{N}_n(0, I)$ and P be an m-dimensional orthogonal projection in $\mathbf{R}^n$. Then we have*

$$Z'PZ \sim \chi^2_m$$

Proof Let

$$P = H_m H'_m$$

where H_m is an orthogonal matrix such that $H'_m H_m = I_m$. Write

$$Z'PZ = (H'_m Z)'(H'_m Z)$$

and observe that $H'_m Z \sim \mathbf{N}_m(0, I)$ to finish the proof. ■

Theorem 10 *Let $Z \sim \mathbf{N}(0, I)$, and let A and B be nonrandom matrices. Then $A'Z$ and $B'Z$ are independent if and only if $A'B = 0$.*

Proof Let

$$C = (A, B)$$

We may assume w.l.o.g. that the matrix C is of full column rank, by throwing away linearly dependent columns, if any, in A and B. Clearly, $C'Z$ is multivariate normal with sub-vectors $A'Z$ and $B'Z$, the covariance of which is zero if and only if $A'B = 0$. The stated result follows from Theorem 6. ■

Corollary 11 *Let $Z \sim \mathbf{N}(0, I)$, and let P and Q are orthogonal projections such that $PQ = 0$. Then $Z'PZ$ and $Z'QZ$ are independent.*

Proof Since $Z'PZ = (PZ)'PZ$ and $Z'QZ = (QZ)'QZ$, $Z'PZ$ and $Z'QZ$ are functions, respectively, of PZ and QZ. It therefore sufficies to show that PZ and QZ are independent, which follows from $PQ = 0$ and Theorem 10. ■

Theorem 12 *Let $Z \sim \mathbf{N}(0, I)$, and let P and Q be orthogonal projections of dimension p and q, respectively. If $PQ = 0$, then*

$$\frac{\frac{Z'PZ}{p}}{\frac{Z'QZ}{q}} \sim F_{p,q}$$

Proof The stated result follows directly from Theorem 9 and Corollary 11, and the definition of F-distribution. ∎

4. An Important Example

We assume throughout this section that X_i's are independent and identically distributed as $\mathbf{N}(\mu, \sigma^2)$. Define

$$\begin{aligned} \overline{X}_n &= \frac{1}{n}\sum_{i=1}^{n} X_i \\ S_n^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \end{aligned}$$

In the following theorem are presented the main results of elementary statistics.

Theorem 13 *We have*

(a) $\overline{X}_n \sim \mathbf{N}\left(\mu, \frac{\sigma^2}{n}\right)$

(b) $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$

(c) $\overline{X}_n$ *and* S_n^2 *are independent*

(d) $\frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} \sim t_{n-1}$

Proof

(a) Let $X = (X_1, \ldots, X_n)'$ and ι be the n-vector of ones. Write $\overline{X}_n = \frac{1}{n}\iota' X$, from which the stated result follows immediately.

(b) Define

$$P_\iota = \frac{\iota\iota'}{n}$$

i.e., the orthogonal projection on the span of ι. Write

$$\begin{aligned}\frac{n-1}{\sigma^2}S_n^2 &= \frac{1}{\sigma^2}X'(I-P_\iota)X \\ &= \left(\frac{X-\mu\iota}{\sigma}\right)'(I-P_\iota)\left(\frac{X-\mu\iota}{\sigma}\right)\end{aligned}$$

Notice that

$$\frac{X-\mu\iota}{\sigma} \sim \mathbf{N}(0,I)$$

and $I-P_\iota$ is an $(n-1)$-dimensional orthogonal projection. Now apply Theorem 9 to get the stated result.

(c) Write

$$\overline{X}_n = \frac{\iota'}{n}P_\iota X \quad \text{and} \quad S_n^2 = \frac{1}{n-1}((I-P_\iota)X)'(I-P_\iota)X$$

The stated result is immediate from the independence of $P_\iota X$ and $(I-P_\iota)X$, which in turn follows from Theorem 10.

(d) It is straightforward from (a) - (c),

$$\frac{\frac{\sqrt{n}(\overline{X}_n-\mu)}{\sigma}}{\sqrt{\frac{\frac{(n-1)S_n^2}{\sigma^2}}{n-1}}} = \frac{\sqrt{n}(\overline{X}_n-\mu)}{S_n}$$

and the definition of t-distribution. ■

5. Exercises

1. Let $X=(X_1,\ldots,X_n)'$ be a random vector with mean $\mu\iota$ and variance Σ, where μ is a scalar, ι is the n-vector of ones and Σ is an $n\times n$ symmetric matrix. We define

$$\overline{X} = \frac{\sum_{i=1}^n X_i}{n} \quad \text{and} \quad S^2 = \frac{\sum_{i=1}^n (X_i-\overline{X})^2}{n-1}$$

Consider the following assumptions:

(A1) X has multivariate normal distribution.

(A2) $\Sigma = \sigma^2 I$

(A3) $\mu = 0$

We claim:

(a) $\overline{X}$ and S^2 are uncorrelated.

(b) $E(\overline{X}) = \mu$

(c) $E(S^2) = \sigma^2$

(d) $\overline{X} \sim N(\mu, \sigma^2/n)$

(e) $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$

(f) $\sqrt{n}(\overline{X} - \mu)/S \sim t_{n-1}$

What assumptions in (A1) – (A3) are needed for each of (a) – (f) to hold. Prove (a) – (f) using the assumptions you specified.

2. Let X and Y have bivariate normal distribution with mean and variance

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

(a) Find a constant α^* such that $Y - \alpha^* X$ is independent of X. Show that $\mathrm{var}(Y - \alpha X) \geq \mathrm{var}(Y - \alpha^* X)$ for any constant α.

(b) Find the conditional distribution of $X + Y$ given $X - Y$.

(c) Obtain $\mathbf{E}(X|X+Y)$.