Lecture 15: Multivariate normal distributions

Normal distributions with singular covariance matrices

Consider an n-dimensional $X \sim N(\mu, \Sigma)$ with a positive definite Σ and a fixed $k \times n$ matrix A that is not of rank k (so k may be larger than n).

The mgf of $Y = AX$ is still equal to

$$M_Y(t) = e^{(A\mu)'t + t'(A\Sigma A')t/2}, \quad t \in \mathscr{R}^k$$

But what is the distribution corresponding to this mgf?

Lemma.

For any $n \times n$ non-negative definite matrix Σ and $\mu \in \mathscr{R}^n$, $e^{\mu' t + t'\Sigma t/2}$ defined for all $t \in \mathscr{R}^n$ is the mgf of an n-dimensional random vector X.

Proof.

From the theory of linear algebra, a non-negative definite matrix Σ of rank $r < n$ satisfies

$$\Sigma = T' \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} T = C'\Lambda C \qquad T = \begin{pmatrix} C \\ D \end{pmatrix}$$

where Λ is an $r \times r$ diagonal matrix whose all diagonal elements are positive, 0 denotes a matrix of 0's of an appropriate order, C is an $r \times n$ matrix of rank r, T is an $n \times n$ matrix satisfying $TT' = T'T = I_n$ (the identity matrix of order n), $CC' = I_r$, $DC' = 0$, $DD' = I_{n-r}$, and $C'C + D'D = I_n$.

Let Y be an r-dimensional random vector $\sim N(C\mu, \Lambda)$ and define

$$X = T' \begin{pmatrix} Y \\ D\mu \end{pmatrix} = C'Y + D'D\mu$$

Since $Y \sim N(C\mu, \Lambda)$, its mgf is $M_Y(s) = e^{(C\mu)'s + s'\Lambda s/2}$, $s \in \mathscr{R}^r$ and the mgf of X is

$$\begin{aligned} M_X(t) &= e^{(D'D\mu)'t} M_Y(Ct)) = e^{(D'D\mu)'t} e^{(C\mu)'(Ct) + (Ct)'\Lambda(Ct)/2} \\ &= e^{\mu'(D'D + C'C)t + t'C'\Lambda Ct/2} = e^{\mu' t + t'\Sigma t/2} \qquad t \in \mathscr{R}^n \end{aligned}$$

This completes the proof.

Definition

For any fixed $n \times n$ non-negative definite matrix Σ and $\mu \in \mathscr{R}^n$, the distribution of an n-dimensional random vector with mgf $e^{\mu' t + t'\Sigma t/2}$ is called normal distribution and denoted by $N(\mu, \Sigma)$.

- If Σ is positive definition, then this definition is the same as the previous definition using the pdf.
- If $X \sim N(\mu, \Sigma)$ and $Y = AX + b$, then $Y \sim N(A\mu, A\Sigma A')$, regardless of whether $A\Sigma A'$ is singular or not.
- If X is multivariate normal, then any sub-vector of X is also normally distributed.
- If n-dimensional $X \sim N(\mu, \Sigma)$ and the rank of Σ is $r < n$, there exists an $r \times n$ matrix C of rank r and $Y = CX \sim N(C\mu, C\Sigma C')$, where $C\Sigma C'$ is a diagonal matrix whose diagonal elements are all positive, and hence Y has an r-dimensional normal pdf and components of Y are independent.
- If n-dimensional $X \sim N(\mu, \Sigma)$ and the rank of Σ is $r < n$, then, from the previous discussion, $X = C'Y + D'D\mu$, where $Y \sim N(C\mu, C\Sigma C')$ and

$$\begin{aligned} E(X) &= C'E(Y) + D'D\mu = (C'C + D'D)\mu = \mu \\ \mathrm{Var}(X) &= C'\mathrm{Var}(Y)C = C'C\Sigma C'C = \Sigma \end{aligned}$$

Thus, μ and Σ in $N(\mu, \Sigma)$ is still the mean and covariance matrix.

Furthermore, any two components of $X \sim N(\mu, \Sigma)$ are independent iff they are uncorrelated.

This can be shown as follows.

Suppose that X_1 and X_2 are the first two components of X and $\text{Cov}(X_1, X_2) = 0$, i.e., the $(1,2)$th and $(2,1)$th elements of Σ are 0.

Let μ_1 and μ_2 be the first two components of μ and σ_1^2 and σ_2^2 be the first and second diagonal elements of Σ, and let $t = (t_1, t_2, 0, ..., 0)$, $t_1 \in \mathscr{R}$, $t_2 \in \mathscr{R}$.

Then the mgf of (X_1, X_2) is

$$M_{(X_1,X_2)}(t_1, t_2) = e^{\mu' t + t'\Sigma t/2} = e^{\mu_1 t_1 + \sigma_1^2 t_1^2/2} e^{\mu_2 t_2 + \sigma_2^2 t_2^2/2} \quad t_1 \in \mathscr{R},\ t_2 \in \mathscr{R}$$

By Theorem M4, X_1 and X_2 are independent.

Theorem.

An n-dimensional random vector $X \sim N(\mu, \Sigma)$ (regardless of whether Σ is singular or not) iff for any n-dimensional constant vector c, $c'X \sim N(c'\mu, c'\Sigma c)$.

Proof.

We treat a degenerated $X = c$ as $N(c, 0)$.

- If $X \sim N(\mu, \Sigma)$, then $M_X(t) = e^{\mu' t + t'\Sigma t/2}$.
  For any $c \in \mathscr{R}^n$, by the properties of mgf, the mgf of $c'X$ is

  $$M_{c'X}(t) = M_X(ct) = e^{\mu'(ct) + (ct)'\Sigma(ct)/2} = e^{(c'\mu)t + (c'\Sigma c)t^2/2} \qquad t \in \mathscr{R}$$

  which is the mgf of $N(c'\mu, c'\Sigma c)$.
  By uniqueness, $c'X \sim N(c'\mu, c'\Sigma c)$.
- If $c'X \sim N(c'\mu, c'\Sigma c)$ for any $c \in \mathscr{R}^n$, then $t'X \sim N(t'\mu, t'\Sigma t)$ for any $t \in \mathscr{R}^n$ and

  $$M_{t'X}(s) = e^{(t'\mu)s + (t'\Sigma t)s^2/2} \qquad s \in \mathscr{R}$$

  Letting $s = 1$, we obtain

  $$M_{t'X}(1) = e^{(t'\mu) + (t'\Sigma t)/2} = E(e^{t'X}) = M_X(t) \qquad t \in \mathscr{R}^n$$

  By uniqueness, $X \sim N(\mu, \Sigma)$.

The condition **any** $c \in \mathscr{R}^n$ is important.

The uniform distribution on $[a,b]\times[c,d]$

We have shown that the two marginal distributions are uniform distributions on intervals $[a,b]$ and $[c,d]$.

For non-zero constants ξ and ζ, is the distribution of $\xi X+\zeta Y$ a uniform distribution on some interval?

If $(e^{bt}-e^{at})/t$ is defined to be $b-a$ when $t=0$ for any constants $a<b$, then

$$\begin{aligned} M_{X,Y}(t,s) &= \int_a^b\int_c^d e^{tx+sy}\frac{1}{(b-a)(d-c)}dxdy \\ &= \frac{(e^{bt}-e^{at})(e^{ds}-e^{cs})}{(b-a)(d-c)ts} \qquad s,t\in\mathscr{R} \end{aligned}$$

and

$$M_{\xi X+\zeta Y}(t)=E(e^{t(\xi X+\zeta Y)})=\frac{(e^{b\xi t}-e^{a\xi t})(e^{d\zeta t}-e^{c\zeta t})}{(b-a)(d-c)\xi\zeta t^2} \qquad t\in\mathscr{R}$$

This is not a mgf of a uniform distribution on an interval $[r,h]$, which is of the form $(e^{ht}-e^{rt})/[t(h-r)]$ for $t\in\mathscr{R}$.

We have shown that if $X \sim N(\mu, \Sigma)$, then any linear function $AX + b$ is normally distributed.

The following result concerns the independence of linear functions of a normally distributed random vector.

Theorem N1.

Let X be an n-dimensional random vector $\sim N(\mu, \Sigma)$ and A be a fixed $k \times n$ matrix, and B be a fixed $l \times n$ matrix. Then, AX and BX are independent iff $A\Sigma B' = 0$.

Proof.

Let

$$Y = \begin{pmatrix} A \\ B \end{pmatrix} X = \begin{pmatrix} AX \\ BX \end{pmatrix}$$

From the properties of the multivariate normal distribution, we know that Y is multivariate normal with covariance matrix

$$\begin{pmatrix} A \\ B \end{pmatrix} \Sigma (A' \;\; B') = \begin{pmatrix} A\Sigma A' & A\Sigma B' \\ B\Sigma A' & B\Sigma B' \end{pmatrix}$$

Hence, AX and BX are uncorrelated iff $A\Sigma B' = 0$ and, thus, the only if part follows since independence implies no correlation.

The proof for the if part is the same as the proof of two uncorrelated components of X are independent: we can show that if $A\Sigma B' = 0$, then the mgf of (AX, BX) is a product of an mgf on $\mathscr{R}^k$ and another mgf on $\mathscr{R}^l$, and then apply Theorem M4.

Theorem N2.

If (X, Y) is a random vector $\sim N(\mu, \Sigma)$ with

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

and if Σ is positive definite, then

$$Y|X \sim N\left(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)$$

It follows from the properties of normal distributions that

$$E(Y|X) = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X - \mu_1), \qquad \mathrm{Var}(Y|X) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

While the conditional mean depends on X, the conditional covariance matrix does not.

Proof.

Consider the transformation

$$U = AX + Y$$

with a fixed matrix A chosen so that U and X are independent.

From Theorem N1, we need U and X to be uncorrelated.

Since

$$\begin{aligned} \mathrm{Cov}(X,U) &= \mathrm{Cov}(X, AX+Y) = \mathrm{Cov}(X, AX) + \mathrm{Cov}(X, Y) \\ &= \mathrm{Cov}(X,X)A' + \Sigma_{12} = \Sigma_{11}A' + \Sigma_{12} \end{aligned}$$

we choose $A = -\Sigma_{21}\Sigma_{11}^{-1}$.

Consider the transformation

$$\begin{pmatrix} V \\ U \end{pmatrix} = \begin{pmatrix} X \\ AX+Y \end{pmatrix} = \begin{pmatrix} I & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}, \qquad \left| \frac{\partial(U,V)}{\partial(X,Y)} \right| = 1$$

Let $f_{(X,Y)}$ be the pdf of (X,Y), $f_{(U,V)}$ be the pdf of (U,V), f_U be the pdf of U and f_V be the pdf of V.

By the transformation formula and the independence of U and $V = X$,

$$f_{(X,Y)}(x,y) = f_{(U,V)}(u,v) = f_U(u)f_V(v) = f_U(y - \Sigma_{21}\Sigma_{11}^{-1}x)f_X(x)$$

Then the pdf of $Y|X$ is

$$\frac{f_{(X,Y)}(x,y)}{f_X(x)} = \frac{f_U(y - \Sigma_{21}\Sigma_{11}^{-1}x)f_X(x)}{f_X(x)} = f_U(y - \Sigma_{21}\Sigma_{11}^{-1}x)$$

Since $U = -\Sigma_{21}\Sigma_{11}^{-1}X + Y$, U is normally distributed.

$$E(U) = -\Sigma_{21}\Sigma_{11}^{-1}E(X) + E(Y) = -\Sigma_{21}\Sigma_{11}^{-1}\mu_1 + \mu_2$$

$$\begin{aligned}
\mathrm{Var}(U) &= \mathrm{Var}(AX + Y) = \mathrm{Var}(AX) + \mathrm{Var}(Y) + 2\mathrm{Cov}(AX, Y) \\
&= A\mathrm{Var}(X)A' + \Sigma_{22} + 2A\mathrm{Cov}(X, Y) \\
&= \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{11}\Sigma_{11}^{-1}\Sigma_{12} + \Sigma_{22} - 2\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \\
&= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}
\end{aligned}$$

Hence, f_U is the pdf of $N(\mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$.

Given $X = x$, $\Sigma_{21}\Sigma_{11}^{-1}x$ is a constant and, hence, $f_U(y - \Sigma_{21}\Sigma_{11}^{-1}x)$ is the pdf of $N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$, considered as a function of y.

Quadratic forms

For a random vector X and a fixed symmetric matrix A, $X'AX$ is called a quadratic function or quadratic form of X.

We now study the distribution of quadratic forms when X is multivariate normal.

Theorem N3.

Let $X \sim N(\mu, I_n)$ and A be a fixed $n \times n$ symmetric matrix. A necessary and sufficient condition for $X'AX$ is chi-square distributed is $A^2 = A$, in which case the degrees of freedom of the chi-square distribution is the rank of A and the noncentrality parameter $\mu' A \mu$.

Proof.

Sufficiency.

If $A^2 = A$, then A is a projection matrix and there exists an $n \times n$ matrix T such that $T'T = TT' = I_n$ and

$$A = T' \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} T = C'C$$

where k is the rank of A and C is the first k rows of T.

Then $X'AX = (CX)'(CX)$ is simply the sum of the squares of CX, the first k components of TX.

Since $TX \sim N(T\mu, TI_nT') = N(T\mu, I_n)$, by definition $X'AX$ has the chi-square distribution with degrees of freedom k and noncentrality parameter $(C\mu)'(C\mu) = \mu'C'C\mu = \mu'A\mu$.

Necessity.
Suppose that $X'AX$ is chi-square with degrees of freedom m and noncentrality parameter $\delta \geq 0$.
Then A must be nonnegative definite and there exists an $n \times n$ matrix T such that $T'T = TT' = I_n$ and

$$A = T' \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} T$$

where Λ is a $k \times k$ diagonal matrix contains k non-zero eigenvalues $0 < \lambda_1 \leq \cdots \leq \lambda_k$.
We still have $TX \sim N(T\mu, I_n)$.
Let $Y_1, ..., Y_k$ be the first k components of TX.
Then Y_i^2's are independent and $Y_i^2 \sim$ chi-square with degree of freedom 1 and noncentrality parameter μ_i^2, where μ_i is the ith

component of μ, and

$$X'AX = \sum_{i=1}^{k} \lambda_i Y_i^2$$

Using the mgf formula for noncentral chi-square distributions, the mgf's of the left and right hand sides are respectively given in the left and right hand sides of the following:

$$\frac{e^{\delta t/(1-2t)}}{(1-2t)^{m/2}} = \prod_{i=1}^{k} \frac{e^{\lambda_i \mu_i^2 t/(1-2\lambda_i t)}}{(1-2\lambda_i t)^{1/2}} \qquad t < 1/2$$

Suppose that $\lambda_k > 1$.
When $t \to (2\lambda_k)^{-1}$, the right hand side of the above equation diverges to ∞ whereas the left hand side of the above equation goes to $e^{\delta(2\lambda_k)^{-1}/(1-\lambda_k^{-1})}/(1-\lambda_k^{-1})^{m/2} < \infty$, which is a contradiction.
Hence $\lambda_k \leq 1$ so that $\lambda_i \leq 1$ for all i.
Suppose that $\lambda_k = \cdots = \lambda_{l+1} = 1 > \lambda_l \geq \cdots \geq \lambda_1 > 0$ for a positive integer $l \leq k$, which implies

$$\frac{e^{\delta t/(1-2t)}}{(1-2t)^{(m-k+l)/2}} = \prod_{i=1}^{l} \frac{e^{\lambda_i \mu_i^2 t/(1-2\lambda_i t)}}{(1-2\lambda_i t)^{1/2}} \qquad t < 1/2$$

When $t \to 1/2$, the left hand side of the above equation diverges to ∞, whereas the right hand side of the above equation converges to

$$\prod_{i=1}^{l} \frac{e^{\lambda_i \mu_i^2 / 2(1-\lambda_i)}}{(1-\lambda_i)^{1/2}}$$

which is a contradiction.
Therefore, we must have $\lambda_1 = \cdots = \lambda_k = 1$, i.e., A is a projection matrix.

Theorem N4 (Cochran's theorem).

Suppose that X is an n-dimensional random vector $\sim N(\mu, I_n)$ and

$$X'X = X'A_1X + \cdots + X'A_kX,$$

where I_n is the $n \times n$ identity matrix and A_i is an $n \times n$ symmetric matrix with rank n_i, $i = 1, ..., k$. A necessary and sufficient condition for

(i) $X'A_iX$ has the noncentral chi-square distribution with degrees of freedom n_i and noncentrality parameter δ_i, $i = 1, ..., k$,

(ii) $X'A_iX$'s are independent,

is $n = n_1 + \cdots + n_k$, in which case $\delta_i = \mu' A_i \mu$ and $\delta_1 + \cdots + \delta_k = \mu'\mu$.

Proof.

Suppose that (i)-(ii) hold.

Then $X'X$ has the chi-square distribution with degrees of freedom $n_1 + \cdots + n_k$ and noncentrality parameter $\delta_1 + \cdots + \delta_k$.

By definition, $X'X$ has the noncentral chi-square distribution with degrees of freedom n and noncentrality parameter $\mu'\mu$.

Then we must have $n = n_1 + \cdots + n_k$ and $\delta_1 + \cdots + \delta_k = \mu'\mu$.

Suppose now that $n = n_1 + \cdots + n_k$.

From the theory of linear algebra, for each i there exists $c_{ij} \in \mathscr{R}^n$, $j = 1, ..., n_i$, such that

$$X'A_iX = \pm(c_{i1}'X)^2 \pm \cdots \pm (c_{in_i}'X)^2$$

Let C be the $n \times n$ matrix whose columns are $c_{11}, ..., c_{1n_1}, ..., c_{k1}, ..., c_{kn_k}$, Then

$$X'X = X'C\Delta C'X$$

with an $n \times n$ diagonal matrix Δ whose diagonal elements are ± 1.

This implies $C\Delta C' = I_n$ and thus C is of full rank and $\Delta = C^{-1}(C')^{-1}$, which is positive definite.

This shows $\Delta = I_n$, which implies $C'C = CC' = I_n$ and

$$X'A_iX = \sum_{j=n_1+\cdots+n_{i-1}+1}^{n_1+\cdots+n_{i-1}+n_i} Y_j^2,$$

where Y_j is the jth component of $Y = C'X \sim N(C'\mu, I_n)$.

Hence Y_j's are independent and $Y_j \sim N(\lambda_j, 1)$, where λ_j is the jth component of $C'\mu$.

This shows that $X'A_iX$, $i = 1,...,k$, are independent and $X'A_iX$ has the chi-square distribution with degrees of freedom n_i and noncentrality parameter $\delta_i = \lambda^2_{n_1+\cdots+n_{i-1}+1} + \cdots + \lambda^2_{n_1+\cdots+n_{i-1}+n_i}$.

Letting $X = \mu$ and $Y = C'X = C'\mu$, we obtain that $\delta_i = \mu' A_i \mu$ and $\delta_1 + \cdots + \delta_k = \mu' CC' \mu = \mu'\mu$.

This completes the proof.

Theorem N5.

Let X be an n-dimensional random vector $\sim N(\mu, I_n)$ and A_1 and A_2 be $n \times n$ projection matrices. Then a necessary and sufficient condition that $X'A_1X$ and $X'A_2X$ are independent is $A_1A_2 = 0$.

Proof.

If $A_1A_2 = 0$, then

$$\begin{aligned}(I_n - A_1 - A_2)^2 &= I_n - A_1 - A_2 - A_1 + A_1^2 + A_2A_1 - A_2 + A_1A_2 + A_2^2 \\ &= I_n - A_1 - A_2,\end{aligned}$$

i.e., $I_n - A_1 - A_2$ is a projection matrix with rank $=$ trace$(I_n - A_1 - A_2)$ $= n - r_1 - r_2$, where $r_i =$ trace(A_i) is the rank of A_i, $i = 1, 2$.

By Cochran's theorem and

$$X'X = X'A_1X + X'A_2X + X'(I_n - A_1 - A_2)X,$$

$X'A_1X$ and $X'A_2X$ are independent.

This proves the sufficiency.

Assume that $X'A_1X$ and $X'A_2X$ are independent.

Since $X'A_iX$ has the noncentral chi-square distribution with degrees of freedom $r_i =$ the rank of A_i and noncentrality parameter $\delta_i = \mu'A_i\mu$, $X'(A_1 + A_2)X$ has the noncentral chi-square distribution with degrees of freedom $r_1 + r_2$ and noncentrality parameter $\delta_1 + \delta_2$.

Consequently, $A_1 + A_2$ is a projection matrix, i.e.,

$$(A_1 + A_2)^2 = A_1 + A_2,$$

which implies

$$A_1 A_2 + A_2 A_1 = 0.$$

Since $A_1^2 = A_1$, we obtain that

$$0 = A_1(A_1 A_2 + A_2 A_1) = A_1 A_2 + A_1 A_2 A_1$$

and

$$0 = A_1(A_1 A_2 + A_2 A_1)A_1 = 2A_1 A_2 A_1,$$

which imply $A_1 A_2 = 0$.

This proves the necessity.