Lecture 22: Multivariate normal distributions (part I)

We now study properties of the multivariate normal distribution defined in the last lecture.

Bivariate normal distributions

First, we want to show that the definition for the general multivariate normal distribution with $n = 2$ is consistent with the early definition of the bivariate normal distribution.

Let $\mu_1 \in \mathscr{R}$, $\mu_2 \in \mathscr{R}$, $\sigma_1 > 0$, $\sigma_2 > 0$, and $-1 < \rho < 1$ be constants, and

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \qquad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Then

$$|\Sigma| = \sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2 = \sigma_1^2\sigma_2^2(1-\rho^2)$$

$$\Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}$$

$$(x-\mu)'\Sigma^{-1}(x-\mu) = \frac{1}{|\Sigma|} \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}' \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}$$

$$= \frac{1}{|\Sigma|}\begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}'\begin{pmatrix} \sigma_2^2(x_1-\mu_1)-\rho\sigma_1\sigma_2(x_2-\mu_2) \\ -\rho\sigma_1\sigma_2(x_1-\mu_1)+\sigma_1^2(x_2-\mu_2) \end{pmatrix}$$

$$= \frac{\sigma_2^2(x_1-\mu_1)^2-2\rho\sigma_1\sigma_2(x_1-\mu_1)(x_2-\mu_2)+\sigma_1^2(x_2-\mu_2)^2}{\sigma_1^2\sigma_2^2(1-\rho^2)}$$

$$= \frac{(x_1-\mu_1)^2}{\sigma_1^2(1-\rho^2)}-\frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2(1-\rho^2)}+\frac{(x_2-\mu_2)^2}{\sigma_2^2(1-\rho^2)}$$

Thus, when $n=2$,

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

$$= \frac{\exp\left(-\frac{(x_1-\mu_1)^2}{2\sigma_1^2(1-\rho^2)}+\frac{\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2(1-\rho^2)}-\frac{(x_2-\mu_2)^2}{2\sigma_2^2(1-\rho^2)}\right)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

which is the same as the bivariate normal pdf defined earlier.

Transformation and marginals

If X is an n-dimensional random vector $\sim N(\mu,\Sigma)$ and $Y=AX+b$, where A is a fixed $k\times n$ matrix of rank $k\leq n$ and b is a fixed k-dimensional vector, then $Y\sim N(A\mu,A\Sigma A')$.

$$= \frac{1}{|\Sigma|}\begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}'\begin{pmatrix} \sigma_2^2(x_1-\mu_1)-\rho\sigma_1\sigma_2(x_2-\mu_2) \\ -\rho\sigma_1\sigma_2(x_1-\mu_1)+\sigma_1^2(x_2-\mu_2) \end{pmatrix}$$

$$= \frac{\sigma_2^2(x_1-\mu_1)^2-2\rho\sigma_1\sigma_2(x_1-\mu_1)(x_2-\mu_2)+\sigma_1^2(x_2-\mu_2)^2}{\sigma_1^2\sigma_2^2(1-\rho^2)}$$

$$= \frac{(x_1-\mu_1)^2}{\sigma_1^2(1-\rho^2)}-\frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2(1-\rho^2)}+\frac{(x_2-\mu_2)^2}{\sigma_2^2(1-\rho^2)}$$

Thus, when $n=2$,

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

$$= \frac{\exp\left(-\frac{(x_1-\mu_1)^2}{2\sigma_1^2(1-\rho^2)}+\frac{\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2(1-\rho^2)}-\frac{(x_2-\mu_2)^2}{2\sigma_2^2(1-\rho^2)}\right)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

which is the same as the bivariate normal pdf defined earlier.

Transformation and marginals

If X is an n-dimensional random vector $\sim N(\mu,\Sigma)$ and $Y=AX+b$, where A is a fixed $k\times n$ matrix of rank $k\le n$ and b is a fixed k-dimensional vector, then $Y\sim N(A\mu,A\Sigma A')$.

This can be proved by using multivariate transformation, or the mgf.

In the last lecture we showed that the mgf of $X \sim N(\mu, \Sigma)$ is

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

From the properties of mgf, $Y = AX + b$ has mgf

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

which is exactly the mgf for $N(A\mu, A\Sigma A')$, noting that $(A\Sigma A')^{-1}$ exists since A has rank $k \le n$ and Σ has rank n.

The result seems still hold even if $A\Sigma A'$ is singular, but we have not defined the normal distribution with a singular covariance matrix.

If we take $A = (I_k \; 0)$, where I_k is the identity matrix of order k and 0 is the $k \times (n-k)$ matrix of all 0's, then AX is exactly the vector containing the first k components of X and, therefore, we have shown that

- If Y is multivariate normal, then any sub-vector of Y is also normally distributed.

Is the converse true? That is, if both X and Y are normal, should the joint pdf of (X, Y) be always normal?

A counter-example

Consider

$$f(x,y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}[1+g(x,y)], \qquad (x,y) \in \mathscr{R}^2$$

where

$$g(x,y) = \begin{cases} xy & -1 < x < 1,\ -1 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Then $f(x,y) \geq 0$. (Is it a pdf?)

$$\begin{aligned} \int_{-\infty}^{\infty} f(x,y)dy &= \frac{1}{2\pi}\left[\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}dy + \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}g(x,y)dy\right] \\ &= \frac{1}{2\pi}e^{-x^2/2}\left[\int_{-\infty}^{\infty} e^{-y^2/2}dy + xI(|x|<1)\int_{-1}^{1} ye^{-y^2/2}dy\right] \\ &= \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \end{aligned}$$

i.e., the marginal of X is $N(0,1)$ (which also shows that $f(x,y)$ is a pdf). Similarly, the pdf of Y is $N(0,1)$.

However, this $f(x,y)$ is certainly not a bivariate normal pdf.

Other joint distributions may not have the property that the marginal distributions are of the same type as the joint distribution.

Uniform distributions

A general n-dimensional uniform distribution on a set $A \subset \mathscr{R}^n$ has pdf

$$f(x) = \begin{cases} 1/C & x \in A \\ 0 & x \notin A \end{cases} \qquad\qquad C = \int_A dx$$

For $n = 1$ and $A =$ an interval, C is the length of the interval.
For $n = 2$, C is the area of the set A.
Consider for example $n = 2$ and A is the rectangle

$$A = [a,b] \times [c,d] = \{(x,y) \in \mathscr{R}^2 : \; a \le x \le b, \; c \le y \le d\}$$

where a, b, c and d are constants.
In this case, the two marginal distributions are uniform distributions on intervals $[a,b]$ and $[c,d]$, since, when $x \in [a,b]$,

$$\int_{-\infty}^{\infty} f(x,y)dy = \frac{1}{C}\int_c^d dy = \frac{1}{(b-a)(d-c)}\int_c^d dy = \frac{1}{b-a}$$

and $\int_{-\infty}^{\infty} f(x,y)dy = 0$ when $x \notin [a,b]$.

However, if A is a disk

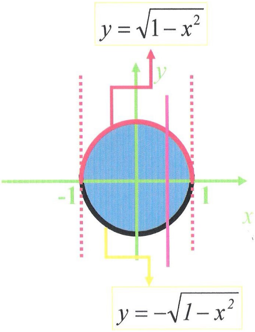
$$A = \{(x,y) \in \mathscr{R}^2 : \; x^2 + y^2 \le 1\}$$

then $C = \pi$ and, when $x \in [-1,1]$,

$$\int_{-\infty}^{\infty} f(x,y)dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy = \frac{2\sqrt{1-x^2}}{\pi}$$

and $\int_{-\infty}^{\infty} f(x,y)dy = 0$ when $x \notin [-1,1]$.

This shows that the marginal distributions are not uniform.



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 $\mu \in \mathscr{R}^n$, the function $e^{\mu' t + t'\Sigma t/2}$ 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

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 $\mu \in \mathscr{R}^n$, the function $e^{\mu' t + t'\Sigma t/2}$ 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 $N(\mu, \Sigma)$.

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, $c'X \sim N(c'\mu, c'\Sigma c)$.

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 $\mathrm{Cov}(X_1,X_2)=0$, 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 $t \in \mathscr{R}$.