The normal distribution

Useful distributions

Francesco Corona

Department of Computer Science
Federal University of Ceará, Fortaleza

The normal distribution

Useful distributions

The normal distribution

Motivation for the normal distribution is found in the central limit theorem

Normal distributions provide an important family of distributions

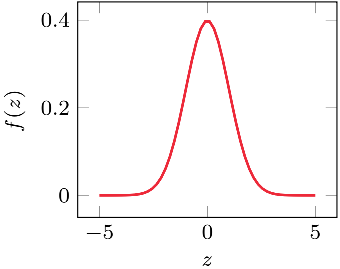
$\rightsquigarrow$ Applications and inference

We first introduce the standard normal distribution

- Then, the general normal distribution

The normal distribution

Consider the integral



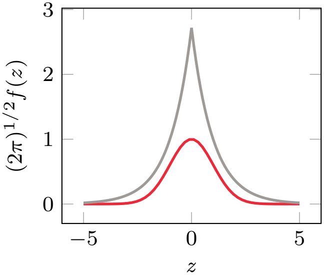
$$I = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\Big(\frac{-z^2}{2}\Big) \mathrm{d}z \quad (1)$$

This integral exists

The normal distribution (cont.)

Consider the relevant part of the integrand

It is a positive continuous function bounded by an integrable function



$$\exp\Big(\frac{-z^2}{2}\Big)$$

That is,

$$0 < \exp\Big(\frac{-z^2}{2}\Big) < \exp\big(-|z|+1\big), \quad -\infty < z < \infty$$

and

$$\int_{-\infty}^{\infty} \exp\big(-|z|+1\big)\mathrm{d}z = 2e$$

The normal distribution (cont.)

To evaluate the integral, we note that $I > 0$

We have that I^2

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{z^2}{2}\Big) \mathrm{d}x \cdot \int_{\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{w^2}{2}\Big) \mathrm{d}w \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\Big(-\frac{z^2+w^2}{2}\Big) \mathrm{d}z \mathrm{d}w \end{aligned}$$

The integral can be computed by changing to polar coordinates

The normal distribution (cont.)

Set $z = r\cos\theta$ and $w = r\sin\theta$

Then,

$$
\begin{aligned}
I^2 &= \frac{1}{2\pi}\int_0^{2\pi} \underbrace{\int_0^{\infty} e^{-r^2/2} r\mathrm{d}r}_{\int x e^{cx^2}\mathrm{d}x = 1/(2c)e^{cx^2} = 1} \mathrm{d}\theta \\
&= \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\theta \\
&= 1
\end{aligned}
$$

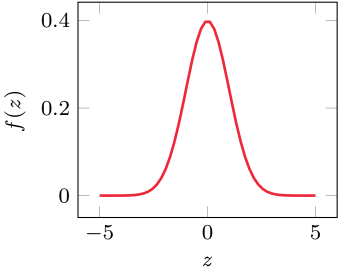
The integrand of Equation (1) is positive on $\mathcal{R}$ and integrates to 1 over $\mathcal{R}$

The normal distribution (cont.)

The integrand is a PDF of a continuous random variable with support $\mathcal{R}$

Denote this random variable Z, Z has the PDF

$$f(z) = \frac{1}{\sqrt{(2\pi)}} \exp\left(\frac{-z^2}{2}\right), \text{ for } -\infty < x < \infty \tag{2}$$



The normal distribution (cont.)

For $t \in \mathcal{R}$, the MGF of Z can be derived by a completion of a square

$$
\begin{aligned}
E[\exp(tZ)] &= \int_{-\infty}^{\infty} \exp(tz) \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2}z^2\Big) \mathrm{d}z \\
&= \exp\Big(\frac{1}{2}t^2\Big) \int_{-\infty}^{\infty} \frac{1}{2\pi} \exp\Big[-\frac{1}{2}(z-t)^2\Big] \mathrm{d}z \qquad (3) \\
&= \exp\Big(\frac{1}{2}t^2\Big) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\Big[-\frac{1}{2}w^2\Big] \mathrm{d}w
\end{aligned}
$$

For the last integral, we made a 1-to-1 change of variable $w = z - t$

The integral in Equation (3) has value 1, by the identity

$$
f(z) = \frac{1}{\sqrt{(2\pi)}} \exp\Big(\frac{-z^2}{2}\Big), \text{ for } -\infty < x < \infty,
$$

The normal distribution (cont.)

Thus, the MGF of Z

$$M_Z(t) = e^{(1/2t^2)}, \text{ for } -\infty < x < \infty \tag{4}$$

The first two derivative of $M_Z(t)$ are

$$M'_Z(t) = te^{(1/2t^2)}$$
$$M''_Z(t) = e^{(1/2t^2)} + t^2 e^{(1/2t^2)}$$

Upon evaluating these derivatives at $t = 0$, the mean and variance of Z

$$\begin{aligned} \rightsquigarrow \quad E(Z) &= 0 \\ \rightsquigarrow \quad \text{Var}(Z) &= 1 \end{aligned} \tag{5}$$

The normal distribution (cont.)

Define the continuous random variable $X = bZ + a$, for $b > 0$

- This is a 1-to-1 transformation

We want to derive the PDF of X

- The inverse of the transformation is $z = b^{-1}(x - a)$
- The Jacobian is $J = b^{-1}$

Because $b > 0$, the PDF of X

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{b} \exp\Big[-\frac{1}{2} \Big(\frac{x-a}{b} \Big)^2 \Big], \text{ for } -\infty < x < \infty$$

The normal distribution (cont.)

By Equation (5), we have

$$\rightsquigarrow \quad E(X) = a$$
$$\rightsquigarrow \quad \mathrm{Var}(X) = b^2$$

The PDF of X can be written using those quantities

- $\rightsquigarrow$ a can be replaced by $\mu = E(X)$
- $\rightsquigarrow$ b^2 can be replaced by $\sigma^2 = \mathrm{Var}(X)$

The normal distribution (cont.)

Definition 1.1

Normal distribution

*A random variable X has a **normal distribution** if its PDF is*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big[-\frac{1}{2}\Big(\frac{x-\mu}{\sigma}\Big)\Big], \quad x \in (-\infty, +\infty) \tag{6}$$

The parameters μ and σ^2 are the mean and the variance of X

We often write that X has a $N(\mu, \sigma^2)$ distribution

The normal distribution (cont.)

Remark

Consider the random variable Z with the PDF

$$f(z) = \frac{1}{\sqrt{(2\pi)}} \exp\Big(\frac{-z^2}{2}\Big), \text{ for } -\infty < x < \infty$$

The RV Z is said to have a $N(0,1)$ distribution

- We call Z a **standard normal** RV

The normal distribution (cont)

Consider the calculation of the MGF of X

We use the relationship

$$X = \underbrace{\sigma}_{b} Z + \underbrace{\mu}_{a}$$

Remember that the MGF of Z

$$M_Z(t) = e^{(1/2t^2)}, \text{ for } -\infty < x < \infty$$

The normal distribution (cont)

We obtain,

$$E\Big[\exp(tX)\Big] = E\Big\{\exp\Big[t(\sigma Z + \mu)\Big]\Big\} = \exp(\mu t) E\Big[\exp(t\sigma Z)\Big]$$
$$= \exp(\mu t)\exp\Big(\frac{1}{2}\sigma^2 t^2\Big) = \exp\Big(\mu t + \frac{1}{2}\sigma^2 t^2\Big)$$

for $-\infty < t < +\infty$

The normal distribution (cont)

Remark

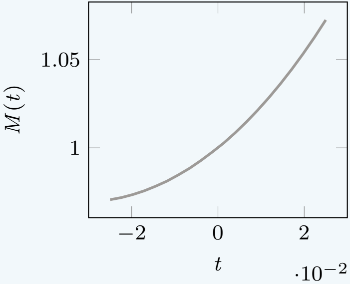
Consider the relationship between Z and X

$\rightsquigarrow$ $X \sim N(\mu, \sigma^2)$ if and only if $Z = (X - \mu)/\sigma \sim N(0, 1)$

The normal distribution (cont.)

Example

Consider the random variable X with the MGF



$$M(t) = e^{2t+32t^2}$$

$$M(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$$

X has a normal distribution with $\mu = 2$ and $\sigma^2 = 64$, $N(2, 54)$

The random variable $Z = (X - 2)/8$ has a $N(0, 1)$ distribution

The normal distribution (cont.)

Example

We can derive all the moments of the standard normal RV using its MGF

- We can do the same for a general normal RV

Let $X \sim N(\mu, \sigma^2)$

Hence, for all nonnegative integers k, by the binomial theorem[1]

$$E(X^k) = E\big[(\sigma Z + \mu)^k\big] = \sum_{j=0}^{k} \binom{k}{j} \sigma^j E(Z^j) \mu^{k-j} \qquad (7)$$

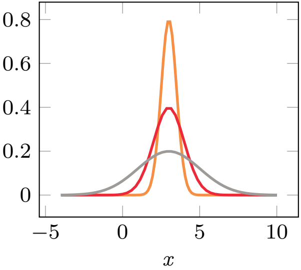
- All the odd moments of Z are zero
- All even moments can be calculated

■

[1] $(x+y)^k = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$

The normal distribution (cont.)

$f(x|\mu = 3, \sigma^2 = 1/4|1|4)$



$$f_X(x) = \frac{1}{\sqrt{2\pi}b} \exp\Big[-\frac{1}{2}\Big(\frac{x-a}{b}\Big)^2\Big]$$

- The graph is symmetric about a vertical axis at $x = \mu$
- The graph has its maximum of $1/(\sigma\sqrt{2\pi})$ at $x = \mu$
- The graph has the x-axis as horizontal asymptote
- The points of inflection are at $x = \mu \pm \sigma$

The normal distribution (cont)

There are many practical applications that involve normal distributions

$\rightsquigarrow$ We need to be able to readily calculate probabilities

Normal PDFs contain a factor of the form $\exp(-s^2)$

- Antiderivatives are not in closed-form
- Numerical integration must be used

The normal distribution (cont)

We can use the relation between normal and standard normal variables

- $\rightsquigarrow$ $X \sim N(\mu, \sigma^2)$ if and only if $Z = (X - \mu)/\sigma \sim N(0, 1)$

Thus, we need only calculate probabilities for standard normal RVs

The normal distribution (cont)

Consider the CDF of a standard normal random variables Z

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\Big(\frac{-w^2}{2}\Big) \mathrm{d}w \tag{8}$$

Let $X \sim N(\mu, \sigma^2)$

Suppose that we want to compute $F_X(x) = P(X \leq x)$ for some x

Then, for $Z = (X - \mu)/\sigma$

$$F_X(x) = P(X \leq x) = P\Big(Z \leq \frac{x-\mu}{\sigma}\Big) = \Phi\Big(\frac{x-u}{\sigma}\Big)$$

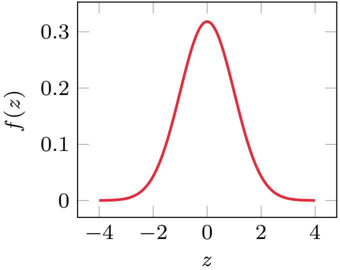
Thus, only integration for $\Phi(z)$ is needed

The normal distribution

Suppose we are interested in the value x_p such that $p = F_X(x_p)$ for some p

Take $z_p = \Phi^{-1}(p)$, then $x_p = \sigma z_p + \mu$

Consider the standard normal density



$$\Phi(z) = \frac{1}{\sqrt{(2\pi)}} \exp\left(\frac{-z^2}{2}\right)$$

The area to the left of z_p is p

- $\Phi(z_p) = p$

We can use an abbreviated table of probability

The normal distribution

Suppose we are interested in $\Phi(-z)$, for some $z > 0$

Since the PDF is symmetric,

$$\rightsquigarrow \quad \Phi(-z) = 1 - \Phi(z) \tag{9}$$

The normal distribution (cont.)

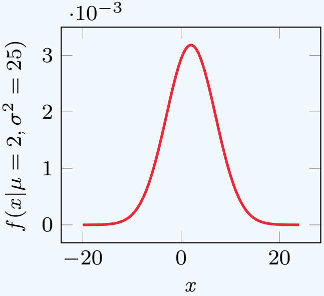
Consider a random variable $X \sim N(a, b)$

- **`pnorm(x,a,b)`**, $P(X \leq x)$
- **`dnorm(x,a,b)`**, the PDF of X at x

The normal distribution (cont.)

Example

Let the random variable $X \sim N(2, 25)$



Using abbreviated tables of probabilities

$$
\begin{aligned}
P(0 < X < 10) &= \Phi\Big(\frac{10-2}{5}\Big) - \Phi\Big(\frac{0-2}{5}\Big) \\
&= \Phi(1.6) - \Phi(-0.4) \\
&= 0.945 - (1 - 0.655) = 0.600
\end{aligned}
$$

The normal distribution (cont.)

$$
\begin{aligned}
P(-8 < X < 1) &= \Phi\Big(\frac{1-2}{5}\Big) - \Phi\Big(\frac{-8-2}{5}\Big) \\
&= \Phi(-0.2) - \Phi(-2) \\
&= (1-0.579) - (1-0.977) = 0.398
\end{aligned}
$$

■

The normal distribution (cont.)

Example

Let the random variable $X \sim N(\mu, \sigma^2)$

$$\begin{aligned} P(\mu - 2\sigma < X < \mu + 2\sigma) &= \Phi\Big(\frac{\mu + 2\sigma - \mu}{\sigma}\Big) - \Phi\Big(\frac{\mu - 2\sigma - \mu}{\sigma}\Big) \\ &= \Phi(2) - \Phi(-2) \\ &= 0.977 - (1 - 0.977) = 0.954 \end{aligned} \quad (10)$$

■

The normal distribution (cont.)

Theorem 1.1

Consider a random variable $X \sim N(\mu, \sigma^2)$ with $\sigma^2 > 0$

The random variable $V = (X - \mu)^2/\sigma^2$ is $\chi^2(1)$

Proof

The random variable $V = W^2$, with $W = (X - \mu)/\sigma \sim N(0, 1)$

The CDF $G(v)$ for V

$$G(v) = P(W^2 \leq v) = P(-\sqrt{v} \leq W \leq \sqrt{v}), \quad \text{for } v \geq 0$$

The normal distribution (cont.)

That is,

$$G(v) = 2\int_0^{\sqrt{v}} \frac{1}{\sqrt{2\pi}} e^{(-w^2/2)} \mathrm{d}w, \text{ for } 0 \leq v$$

Moreover,

$$G(v) = 0, \text{ for } v < 0$$

Consider a change of the integration variable, $w = \sqrt{y}$

$$\rightsquigarrow \quad G(v) = \int_0^{v} \frac{1}{\sqrt{2\pi}\sqrt{y}} e^{(-y/2)} \mathrm{d}y, \text{ for } 0 \leq v$$

The normal distribution (cont.)

Hence, the PDF $g(v) = G'(v)$ of random variable V of the continuous type

$$g(v) = \begin{cases} \dfrac{1}{\sqrt{\pi}\sqrt{2}} v^{(1/2-1)} e^{(-v/2)}, & 0 < v < \infty \\ 0, & \text{elsewhere} \end{cases}$$

Since $g(v)$ is a PDF,

$$\int_0^{\infty} g(v)\mathrm{d}v = 1,$$

Then, it must be $\Gamma(1/2) = \sqrt{\pi}$ and therefore $V \sim \chi^2(1)$

■

The normal distribution (cont.)

A main property of the normal distribution is additivity under independence

Theorem 1.2

Let $X_1, \dots, X_n$ be independent random variables

Suppose that $X_i \sim N(\mu_i, \sigma_i^2)$, for $i = 1, \dots, n$,

Let $Y = \sum_{i=1}^n a_i X_i$, for some constants $a_1, \dots, a_n$

Then, the distribution of Y is $N(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2)$

Proof

For $t \in \mathcal{R}$, the MGF of Y

$$
\begin{aligned}
M_Y(t) &= \prod_{i=1}^n \exp\left[t a_i \mu_i + (1/2) t^2 a_i^2 \sigma_i^2\right] \\
&= \exp\left[t \sum_{i=1}^n a_i \mu_i + (1/2) t^2 \sum_{i=1}^n a_i^2 \sigma_i^2\right]
\end{aligned}
$$

This is the MGF of a $N(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2)$ distribution $\blacksquare$

The normal distribution (cont.)

Let $X_1, X_2, \ldots, X_n$ be random sample from a $N(\mu, \sigma^2)$

A corollary gives the distribution of the sample mean

$$\rightsquigarrow \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Corollary

Let $X_1, \ldots, X_n$ be IID random variables with common $N(\mu, \sigma^2)$ distribution

Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Then, $\overline{X}$ *has a* $N(\mu, \sigma^2/n)$ *distribution*

Proof

In Theorem 1.2, take $a_i = (1/n)$, $\mu_i = \mu$ and $\sigma_i^2 = \sigma^2$, for $i = 1, 2, \ldots, n$

$\blacksquare$

The multivariate normal distribution

Useful distributions

The multivariate normal distribution

The multivariate normal distribution for a n-dimensional random vector

- Examples for the bivariate case, $n = 2$

The derivation is simplified by first discussing the standard case

- Then, we proceed with the general cases

The multivariate normal distribution (cont.)

Consider the random vector $\mathbf{Z} = (Z_1, \ldots, Z_n)'$

- $Z_1, \ldots, Z_n$ are IID $N(0, 1)$ random variables

Then, the density of $\mathbf{Z}$

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{z}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\Big\{ -\frac{1}{2} z_i^2 \Big\} = \Big(\frac{1}{2\pi} \Big)^{n/2} \exp\Big\{ -\frac{1}{2} \sum_{i=1}^{n} z_i^2 \Big\} \\ &= \Big(\frac{1}{2\pi} \Big)^{n/2} \exp\Big\{ -\frac{1}{2} \mathbf{z}'\mathbf{z} \Big\}, \text{ for } \mathbf{z} \in \mathcal{R}^n \end{aligned} \quad (11)$$

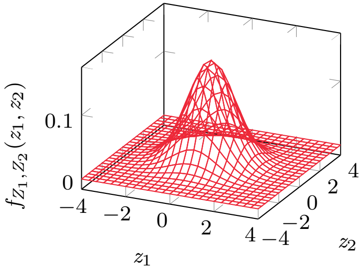
The multivariate normal distribution (cont.)

Consider the random vector $\mathbf{Z} = (Z_1, Z_2)'$

Since $n = 2$, we have

$$f_{\mathbf{Z}}(\mathbf{z}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\Big\{ -\frac{1}{2} z_i^2 \Big\} = \Big(\frac{1}{2\pi}\Big)^{n/2} \exp\Big\{ -\frac{1}{2} \sum_{i=1}^{n} z_i^2 \Big\},$$

$$= 1/(2\pi) \exp\Big\{ -1/2 \sum_{i=1}^{2} z_i^2 \Big\} = 1/(2\pi) \exp\Big\{ -1/2 (z_1^2 + z_2^2) \Big\}$$



The multivariate normal distribution (cont.)

The Z_is have mean zero, unit variance and they are uncorrelated

Then, the mean and covariance matrix of $\mathbf{Z}$

$$\begin{aligned} E(\mathbf{Z}) &= \mathbf{0} \\ \text{Cov}(\mathbf{Z}) &= \mathbf{I}_n \end{aligned} \tag{12}$$

- $\mathbf{I}_n$ indicates an identity matrix of order n

The multivariate normal distribution

The MGF of each of the Z_is evaluated at t_i is $e^{(t_i^2/2)}$

As the Z_is are independent, the MGF of $\mathbf{Z}$

$$
\begin{aligned}
M_{\mathbf{Z}}(\mathbf{t}) &= E\Big[\prod_{i=1}^{n} \exp\{t_i Z_i\}\Big] = \prod_{i=1}^{n} E[\exp\{t_i Z_i\}] \\
&= \exp\Big\{\frac{1}{2}\sum_{i=1}^{n} t_i^2\Big\} = \exp\Big\{\frac{1}{2}\mathbf{t}'\mathbf{t}\Big\}, \text{for all } \mathbf{t} \in \mathcal{R}^n
\end{aligned} \tag{13}
$$

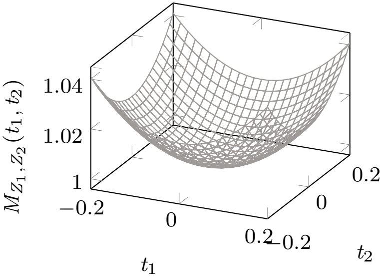
The multivariate normal distribution

Consider the random vector $\mathbf{Z} = (Z_1, Z_2)'$

Since $n = 2$, we have

$$M_{\mathbf{Z}}(\mathbf{t}) = E\Big[\prod_{i=1}^{n} \exp\{t_i Z_i\}\Big] = \prod_{i=1}^{n} E[\exp\{t_i Z_i\}]$$

$$= \exp\Big\{\frac{1}{2}\sum_{i=1}^{2} t_i^2\Big\} = \exp\{1/2(t_1^2 + t_2^2)\}, \text{ for all } (t_1, t_2) \in \mathcal{R}^2$$



The multivariate normal distribution (cont.)

$\mathbf{Z}$ is said to have a **multivariate normal distribution**

$\mathbf{Z}$ has a $N_n(\mathbf{0}, \mathbf{I}_n)$ distribution, mean $\mathbf{0}$, covariance matrix $\mathbf{I}_n$

The multivariate normal distribution (cont.)

Let $\mathbf{\Sigma}$ be a $n \times n$, symmetric and positive-definite matrix $(\mathbf{z}'\mathbf{\Sigma}\mathbf{z} > 0, \forall \mathbf{z})$

From linear algebra, we can always decompose $\mathbf{\Sigma}$

$$\mathbf{\Sigma} = \mathbf{\Gamma}'\mathbf{\Lambda}\mathbf{\Gamma} \tag{14}$$

$\mathbf{\Lambda}$ is the diagonal matrix $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

$\rightsquigarrow$ $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ are the eigenvalues of $\mathbf{\Sigma}$

The columns of $\mathbf{\Gamma}'$, $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$, are the corresponding eigenvectors

$$\mathbf{\Gamma}' = \begin{bmatrix} v_{1,1} & v_{2,1} & \cdots & v_{n,1} \\ v_{1,2} & v_{2,2} & \cdots & v_{n,2} \\ \vdots & \vdots & & \vdots \\ v_{1,n} & v_{2,n} & \cdots & v_{n,n} \end{bmatrix}$$

The multivariate normal distribution (cont.)

$$\mathbf{\Sigma} = \mathbf{\Gamma}'\mathbf{\Lambda}\mathbf{\Gamma}$$

The factorisation is called the **spectral decomposition** of $\mathbf{\Sigma}$

- Matrix $\mathbf{\Gamma}$ is orthogonal ($\mathbf{\Gamma}^{-1} = \mathbf{\Gamma}'$, thus $\mathbf{\Gamma}\mathbf{\Gamma}' = \mathbf{I}$)

The multivariate normal distribution (cont.)

Remark

Eigenvalues and eigenvectors

A $(n \times n)$ matrix $\mathbf{A}$ can be used to transform a n-vector $\mathbf{x}$ to another one $\mathbf{y}$

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

Consider a scalar variable λ_i and a particular value of $\mathbf{x}$ $(= \mathbf{e}_i)$ such that

$$\mathbf{y}_{\mathbf{e}_i} = \mathbf{A}\mathbf{e}_i = \lambda_i \mathbf{e}_i$$

- λ_i is often called an **eigenvalue** of $\mathbf{A}$
- $\mathbf{e}_i$ is the corresponding **eigenvector**

The multivariate normal distribution (cont.)

$$\mathbf{A}\mathbf{e}_i = \lambda_i \mathbf{e}_i$$

Every $(n \times n)$ matrix has n eigenvalues and up to n eigenvectors

They can be jointly specified

$$\begin{aligned}
(\lambda_1 \mathbf{I}_n - \mathbf{A})\mathbf{e}_1 &= \mathbf{0} \\
(\lambda_2 \mathbf{I}_n - \mathbf{A})\mathbf{e}_2 &= \mathbf{0} \\
&\cdots \\
(\lambda_n \mathbf{I}_n - \mathbf{A})\mathbf{e}_n &= \mathbf{0}
\end{aligned}$$

If $(\lambda_i \mathbf{I}_n - \mathbf{A})\mathbf{e}_i = \mathbf{0}$ is satisfied, then so is $(\lambda_i \mathbf{I}_n - \mathbf{A})\alpha\mathbf{e}_i = \mathbf{0} \quad (i = 1, \dots, n)$

$\rightsquigarrow$ Eigenvectors are specified up to any multiplicative constant α

The multivariate normal distribution (cont.)

The set of n vector equations can have non-trivial solutions

The n values of λ_i need be solutions to the scalar equation

$$|\lambda \mathbf{I}_n - \mathbf{A}| := \Delta(\lambda) = 0$$

⇝ $(\lambda \mathbf{I}_n - \mathbf{A})$ is called the **characteristic matrix** of $\mathbf{A}$

⇝ $\Delta(\lambda)$ is the **characteristic polynomial** of $\mathbf{A}$

The determinant gives a n-degree polynomial in λ

- It can be factored as a product of n binomials

$$\begin{aligned}\Delta(\lambda) &= \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0\\ &= (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n)\\ &= 0\end{aligned}$$

- Each of the binomial roots is an eigenvalue of $\mathbf{A}$

The multivariate normal distribution (cont.)

If the roots are all distinct, there will be n independent eigenvectors

$\rightsquigarrow$ If the roots are repeated, there may be fewer eigenvectors

For a distinct eigenvalue λ_i, the eigenvector $\mathbf{e}_i$ is contained in $\text{Adj}(\lambda_i \mathbf{I}_n - \mathbf{A})$

$$\text{Adj}(\lambda_i \mathbf{I}_n - \mathbf{A}) = \begin{bmatrix} \alpha_1 \mathbf{e}_i & \alpha_2 \mathbf{e}_i & \cdots & \alpha_n \mathbf{e}_i \end{bmatrix}$$

Since α_j are arbitrary constants, any non-zero column represents $\mathbf{e}_i$

$\text{Adj}(\lambda_i \mathbf{I}_n - \mathbf{A})$ is computed for each root $(i = 1, \dots, n)$

- A single eigenvector is chosen from each evaluation

The multivariate normal distribution (cont.)

Together, the n eigenvectors form the columns of the **modal matrix** $\mathbf{E}$

$$\mathbf{E} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix}$$

Here, the eigenvectors are scaled so that $|\mathbf{e}_i| = 1$ $(i = 1, \ldots, n)$

The multivariate normal distribution (cont.)

The modal matrix can be used to diagonalise $\mathbf{A}$

To transform $\mathbf{A}$ into a diagonal matrix (of eigenvalues)

$$\begin{bmatrix} \mathbf{A}\mathbf{e}_1 & = & \lambda_1 \mathbf{e}_1 \\ \mathbf{A}\mathbf{e}_2 & = & \lambda_2 \mathbf{e}_2 \\ \vdots & & \vdots \\ \mathbf{A}\mathbf{e}_n & = & \lambda_n \mathbf{e}_n \end{bmatrix}, \text{ or } \mathbf{AE} = \mathbf{E}\mathbf{\Lambda}$$

$\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

The multivariate normal distribution (cont.)

$\mathbf{E}$ is non-singular because the $\mathbf{e}_i$ are linearly independent

$$\mathbf{A} = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^{-1}$$

From the diagonal matrix of eigenvalues to the original square matrix

The inverse transformation

$$\boldsymbol{\Lambda} = \mathbf{E}^{-1}\mathbf{A}\mathbf{E}$$

The multivariate normal distribution (cont.)

If $\mathbf{A}$ is a real, symmetric matrix, its eigenvalues are real

- The eigenvectors are orthogonal to each other

There are n eigenvectors, even with repeated roots

The eigenvectors corresponding to distinct eigenvalues

$$\text{Adj}(\lambda_i \mathbf{I}_n - \mathbf{A}) = \begin{bmatrix} \alpha_1 \mathbf{e}_i & \alpha_2 \mathbf{e}_i & \cdots & \alpha_n \mathbf{e}_i \end{bmatrix}$$

The eigenvectors of eigenvalues λ_i with multiplicity m

$$\frac{d^{m-1}}{d\lambda^{m-1}} \big[\text{Adj}(\lambda_i \mathbf{I}_n - \mathbf{A})\big]\Big|_{\lambda=\lambda_i} = \begin{bmatrix} \cdots & \mathbf{e}_{i_1} & \mathbf{e}_{i_2} & \cdots & \mathbf{e}_{i_m} & \cdots \end{bmatrix}$$

The multivariate normal distribution (cont.)

Remark

An alternative way to write the spectral decomposition

$$\mathbf{\Sigma} = \mathbf{\Gamma}'\mathbf{\Lambda}\mathbf{\Gamma} = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i' \tag{15}$$

$$\underbrace{\begin{bmatrix} v_{1,1} & v_{2,1} & \cdots & v_{n,1} \\ v_{1,2} & v_{2,2} & \cdots & v_{n,2} \\ \vdots & \vdots & & \vdots \\ v_{1,n} & v_{2,n} & \cdots & v_{n,n} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}}_{\lambda_2 \mathbf{v}_2} \underbrace{\begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,n} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,n} \\ \vdots & \vdots & & \vdots \\ v_{n,1} & v_{n,2} & \cdots & v_{n,n} \end{bmatrix}}_{\mathbf{v}_2'}$$

The multivariate normal distribution (cont.)

Because the $\lambda_i s$ are non-negative, we can define the diagonal matrix

$$\mathbf{\Lambda}^{1/2} = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_n})$$

$$= \begin{bmatrix} {\lambda_1}^{1/2} & 0 & \cdots & 0 \\ 0 & {\lambda_2}^{1/2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & {\lambda_n}^{1/2} \end{bmatrix}$$

Then, the orthogonality of $\mathbf{\Gamma}$ implies

$$\mathbf{\Sigma} = \mathbf{\Gamma}' \underbrace{\mathbf{\Lambda}^{1/2} \underbrace{\mathbf{\Gamma}\mathbf{\Gamma}'}_{\mathbf{I}} \mathbf{\Lambda}^{1/2}}_{\mathbf{\Lambda}} \mathbf{\Gamma}$$

The multivariate normal distribution (cont.)

$$\mathbf{\Sigma} = \underbrace{\mathbf{\Gamma}' \, \mathbf{\Lambda}^{1/2} \, \underbrace{\mathbf{\Gamma}\mathbf{\Gamma}'}_{\mathbf{I}} \, \mathbf{\Lambda}^{1/2}}_{\mathbf{\Lambda}} \, \mathbf{\Gamma}$$

Define the square root of the positive semi-definite matrix $\mathbf{\Sigma}$

$$\mathbf{\Sigma}^{1/2} = \mathbf{\Gamma}' \mathbf{\Lambda}^{1/2} \mathbf{\Gamma} \tag{16}$$

- $\mathbf{\Sigma}^{1/2}$ is symmetric and positive semi-definite

Suppose $\mathbf{\Sigma}$ is positive definite and all its eigenvalues are strictly positive

Then, we have

$$(\mathbf{\Sigma}^{1/2})^{-1} = \mathbf{\Gamma}' \mathbf{\Lambda}^{-1} \mathbf{\Gamma} = \mathbf{\Sigma}^{-1/2} \tag{17}$$

The multivariate normal distribution (cont.)

Let $\mathbf{Z}$ have a $N_n(\mathbf{0}, \mathbf{I}_n)$ distribution

Let $\mathbf{\Sigma}$ be a positive semi-definite symmetric matrix

Let $\boldsymbol{\mu}$ be a $n \times 1$ vector of constants

Define the random vector $\mathbf{X}$

$$\mathbf{X} = \mathbf{\Sigma}^{1/2}\mathbf{Z} + \boldsymbol{\mu} \tag{18}$$

By Equation (12), we have

$$\begin{aligned} E[\mathbf{X}] &= \boldsymbol{\mu} \\ \text{Cov}[\mathbf{X}] &= \mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{1/2} = \mathbf{\Sigma} \end{aligned} \tag{19}$$

The multivariate normal distribution (cont.)

The MGF of $\mathbf{X}$

$$\begin{aligned} M_{\mathbf{X}}(\mathbf{t}) = E\Big[\exp\big\{\mathbf{t}'\mathbf{X}\big\}\Big] &= E\Big[\exp\big\{\mathbf{t}'\boldsymbol{\Sigma}^{1/2}\mathbf{Z} + \mathbf{t}'\boldsymbol{\mu}\big\}\Big] \\ &= \exp\big(\mathbf{t}'\boldsymbol{\mu}\big)E\Big\{\exp\Big[\big(\boldsymbol{\Sigma}^{1/2}\mathbf{t}\big)'\mathbf{Z}\Big]\Big\} \\ &= \exp\big(\mathbf{t}'\boldsymbol{\mu}\big)\exp\Big[(1/2)\big(\boldsymbol{\Sigma}^{1/2}\mathbf{t}\big)'\boldsymbol{\Sigma}^{1/2}\mathbf{t}\Big] \\ &= \exp\big(\mathbf{t}'\boldsymbol{\mu}\big)\exp\Big[(1/2)\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\Big] \end{aligned} \tag{20}$$

The multivariate normal distribution (cont.)

Definition

Multivariate normal

A n-dimensional random vector $\mathbf{X}$ has a ***multivariate normal distribution*** *if its MGF is*

$$M_{\mathbf{X}}(\mathbf{t}) = \exp\left[\mathbf{t}'\boldsymbol{\mu} + (1/2)\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\right] \qquad (21)$$

for all $\mathbf{t} \in \mathcal{R}^n$

- $\boldsymbol{\Sigma}$ *is a symmetric, positive semi-definite matrix*
- $\boldsymbol{\mu} \in \mathcal{R}^n$

We say that $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution

The multivariate normal distribution (cont.)

The definition is for positive semi-definite positive matrices $\mathbf{\Sigma}$

- $\mathbf{\Sigma}$ is usually positive definite

$\rightsquigarrow$ We can get the density of $\mathbf{X}$

If $\mathbf{\Sigma}$ is positive definite, then so is $\mathbf{\Sigma}^{1/2}$

- Its inverse is $(\mathbf{\Sigma}^{1/2})^{-1} = \mathbf{\Gamma}'\mathbf{\Lambda}^{-1}\mathbf{\Gamma} = \mathbf{\Sigma}^{-1/2}$

The multivariate normal distribution (cont.)

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

The transformation between $\mathbf{X}$ and $\mathbf{Z}$ is 1-to-1

The inverse transformation

$$\mathbf{Z} = \mathbf{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$$

The Jacobian

$$\left|\mathbf{\Sigma}^{-1/2}\right| = \left|\mathbf{\Sigma}\right|^{-1/2}$$

The multivariate normal distribution (cont.)

Hence, upon simplification, the PDF of $\mathbf{X}$

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\Big[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\Big], \text{ for } \mathbf{x} \in \mathcal{R}^n \quad (22)$$

The multivariate normal distribution (cont.)

Theorem 2.1

Suppose $\mathbf{X}$ *has a* $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *distribution*

Let $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$, *where* $\mathbf{A}$ *is a* $m \times n$ *matrix and* $\mathbf{b} \in \mathcal{R}^m$

Then, $\mathbf{Y}$ *has a* $N_m(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$ *distribution*

The multivariate normal distribution (cont.)

Proof

From $M_{\mathbf{X}}(\mathbf{t}) \exp \big[\mathbf{t}'\boldsymbol{\mu} + (1/2)\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\big]$, for all $\mathbf{t} \in \mathcal{R}^m$

The MGF of $\mathbf{Y}$

$$
\begin{aligned}
M_{\mathbf{Y}}(\mathbf{t}) &= E\big[\exp\big(\mathbf{t}'\mathbf{Y}\big)\big] \\
&= E\big\{\exp\big[\mathbf{t}'(\mathbf{A}\mathbf{X} + \mathbf{b})\big]\big\} \\
&= \exp(\mathbf{t}'\mathbf{b}) E\big\{\exp\big[\big(\mathbf{A}'\mathbf{t}\big)'\mathbf{X}\big]\big\} \\
&= \exp(\mathbf{t}'\mathbf{b}) \exp\big[\big(\mathbf{A}'\mathbf{t}\big)'\boldsymbol{\mu} + \big(1/2\big)\big(\mathbf{A}'\mathbf{t}\big)'\boldsymbol{\Sigma}\big(\mathbf{A}'\mathbf{t}\big)\big] \\
&= \exp\big[\mathbf{t}'\big(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}\big) + \big(1/2\big)\mathbf{t}'\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\mathbf{t}\big]
\end{aligned}
$$

This is the MGF of a $N_m(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$ distribution

■

The multivariate normal distribution (cont.)

A corollary gives us the marginal distributions of a multivariate normal RV

Let $\mathbf{X}_1$ be any sub-vector of $\mathbf{X}$ of dimension $m < n$

We can always rearrange means and correlations

There is no loss in writing

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \tag{23}$$

$\mathbf{X}_2$ is of dimension $p = n - m$

The multivariate normal distribution (cont.)

In the same way, we partition mean and covariance matrix of $\mathbf{X}$

$$\begin{aligned} \boldsymbol{\mu} &= \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \\ \boldsymbol{\Sigma} &= \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \end{aligned} \tag{24}$$

- $\boldsymbol{\Sigma}_{11}$ is the covariance matrix of $\mathbf{X}_1$
- $\boldsymbol{\Sigma}_{12}$ contains all covariances between the components of $\mathbf{X}_1$ and $\mathbf{X}_2$

The multivariate normal distribution (cont.)

We define $\mathbf{A}$ to be the matrix

$$\mathbf{A} = \Big[[\mathbf{I}_m] \Big| [\mathbf{0}_{mp}] \Big] = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & \cdots & 0 \end{bmatrix}$$

- $\mathbf{I}_m$ indicates a $m \times m$ identity matrix
- $\mathbf{0}_{mp}$ indicates a $m \times p$ matrix of zeros

Then, $\mathbf{X}_1 = \mathbf{A}\mathbf{X}$

The multivariate normal distribution (cont.)

We apply Theorem 2.1 to the transformation $\mathbf{X}_1 = \mathbf{A}\mathbf{X}$

We have the following corollary

Corollary 2.1

Suppose $\mathbf{X}$ *has a* $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *distribution, partitioned as in Equation (23-24)*

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$$
$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

Then, $\mathbf{X}_1$ *has a* $N_m(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ *distribution*

This is a useful result

The corollary shows that any marginal distribution of $\mathbf{X}$ is also normal

- $\leadsto$ Mean and covariance matrix are those from the partial vector

The multivariate normal distribution (cont.)

Example

Consider the multivariate normal in the case when $n = 2$

- $\rightsquigarrow$ The bivariate normal

We use common notation (X, Y) rather than (X_1, X_2)

Suppose that $(X, Y) \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$
$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_2^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \quad (25)$$

- $\rightsquigarrow$ μ_1 and σ_1^2 are the mean and variance of X
- $\rightsquigarrow$ μ_2 and σ_2^2 are the mean and variance of Y

$\sigma_{12} = \rho(\sigma_1 \sigma_2)$ is the covariance between X and Y

- $\rightsquigarrow$ ρ is the correlation coefficient between X and Y

The multivariate normal distribution (cont.)

Substituting $\rho\sigma_1\sigma_2$ for σ_{12} in $\mathbf{\Sigma}$

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_2^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 \end{bmatrix}$$

The determinant of $\mathbf{\Sigma}$ is $\sigma_1^2\sigma_2^2(1-\rho^2)$ (remember that $\rho^2 \leq 1$)

- Suppose that $\rho^2 < 1$, so that $\mathbf{\Sigma}$ is invertible

The inverse of $\mathbf{\Sigma}$

$$\mathbf{\Sigma}^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_2\sigma_1 & \sigma_1^2 \end{bmatrix} \qquad (26)$$

The multivariate normal distribution (cont.)

From this expression, the PDF of (X, Y)

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{(-q/2)}, \quad -\infty < x, y < \infty \tag{27}$$

with

$$q = \frac{1}{1-\rho^2}\Big[\Big(\frac{x-\mu_1}{\sigma_1}\Big)^2 - 2\rho\Big(\frac{x-\mu_1}{\sigma_1}\Big)\Big(\frac{x-\mu_2}{\sigma_2}\Big) + \Big(\frac{y-\mu_2}{\sigma^2}\Big)^2\Big] \tag{28}$$

The multivariate normal distribution (cont.)

In general, if X and Y are independent RVs, correlation coefficient is zero

If they are normal, then $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$

- (by Corollary 2.1)

The multivariate normal distribution (cont.)

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{(-q/2)}, \quad -\infty < x, y < \infty$$

$$q = \Big[\Big(\frac{x-\mu_1}{\sigma_1}\Big)^2 + \Big(\frac{y-\mu_2}{\sigma^2}\Big)^2\Big]$$

From the joint PDF of (X, Y), let the correlation coefficient be zero

- X and Y are independent

For the bivariate normal, independence corresponds to $\rho = 0$

The generalisation to the multivariate case also holds true

The multivariate normal distribution (cont.)

If two random variables are independent their covariance is zero

- The converse is not necessarily true

Yet, it can be shown that this is true for the multivariate normal

Theorem 2.2

Suppose that $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Suppose the partitioning

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

Then, $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *are independent if and only if* $\boldsymbol{\Sigma}_{12} = \mathbf{0}$

The multivariate normal distribution (cont.)

Proof

Note that $\mathbf{\Sigma}_{12} = \mathbf{\Sigma}'_{21}$

The joint MGF of $\mathbf{X}_1$ and $\mathbf{X}_2$

$$M_{\mathbf{x}_1,\mathbf{x}_2}(\mathbf{t}_1, \mathbf{t}_2) =$$

$$\exp\Big\{\mathbf{t}'_1\boldsymbol{\mu}_1 + \mathbf{t}'_2\boldsymbol{\mu}_2 + \frac{1}{2}\Big(\mathbf{t}'_1\mathbf{\Sigma}_{11}\mathbf{t}_1 + \mathbf{t}'_2\mathbf{\Sigma}_{22}\mathbf{t}_2 + \mathbf{t}'_2\mathbf{\Sigma}_{21}\mathbf{t}_1 + \mathbf{t}'_1\mathbf{\Sigma}_{12}\mathbf{t}_2\Big)\Big\} \qquad (29)$$

We used

$$\mathbf{t}' = (\mathbf{t}'_1, \mathbf{t}'_2)$$

The multivariate normal distribution (cont.)

By Corollary (2.1),

- $\mathbf{X}_1 \sim N_m(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$
- $\mathbf{X}_2 \sim N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$

The product of their marginal MGFs

$$M_{\mathbf{x}_1}(\mathbf{t}_1)M_{\mathbf{x}_2}(\mathbf{t}_2) = \exp\left\{\mathbf{t}_1'\boldsymbol{\mu}_1 + \mathbf{t}_2'\boldsymbol{\mu}_2 + \frac{1}{2}\Big(\mathbf{t}_1'\boldsymbol{\Sigma}_{11}\mathbf{t}_1 + \mathbf{t}_2'\boldsymbol{\Sigma}_{22}\mathbf{t}_2\Big)\right\} \quad (30)$$

For $\mathbf{X}_1$ and $\mathbf{X}_2$ be independent, Equation (29) and (30) must be identical

The multivariate normal distribution (cont.)

$$\exp\Big\{\mathbf{t}_1'\boldsymbol{\mu}_1 + \mathbf{t}_2'\boldsymbol{\mu}_2 + \frac{1}{2}\Big(\mathbf{t}_1'\boldsymbol{\Sigma}_{11}\mathbf{t}_1 + \mathbf{t}_2'\boldsymbol{\Sigma}_{22}\mathbf{t}_2 + \mathbf{t}_2'\boldsymbol{\Sigma}_{21}\mathbf{t}_1 + \mathbf{t}_1'\boldsymbol{\Sigma}_{12}\mathbf{t}_2\Big)\Big\}$$

$$\exp\Big\{\mathbf{t}_1'\boldsymbol{\mu}_1 + \mathbf{t}_2'\boldsymbol{\mu}_2 + \frac{1}{2}\Big(\mathbf{t}_1'\boldsymbol{\Sigma}_{11}\mathbf{t}_1 + \mathbf{t}_2'\boldsymbol{\Sigma}_{22}\mathbf{t}_2\Big)\Big\}$$

Independence of $\mathbf{X}_1$ and $\mathbf{X}_2$ is verified when $\boldsymbol{\Sigma}_{12} = \mathbf{0}'$ and hence $\boldsymbol{\Sigma}_{21} = \mathbf{0}$

- $\rightsquigarrow$ The covariances between their components are all 0

■

The multivariate normal distribution (cont.)

Theorem

Suppose $\mathbf{X}$ *has a* $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *distribution, partitioned as in Equation (23-24)*

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

Assume that $\boldsymbol{\Sigma}$ *is positive definite*

Then, the conditional distribution of $\mathbf{X}_1 | \mathbf{X}_2$

$$N_m \Big[\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{X}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \Big]$$

The multivariate normal distribution (cont.)

Proof

Consider the joint distribution of $\mathbf{W} = \mathbf{X}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{X}_2$ and $\mathbf{X}_2$

The joint distribution can be obtained from the transformation

$$\begin{bmatrix} \mathbf{W} \\ \mathbf{X}_2 \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{I}_m & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$$

This is a linear transformation

The joint distribution is multivariate normal (Theorem 2.1)

The multivariate normal distribution (cont.)

The means

- $E(\mathbf{W}) = \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2$
- $E(\mathbf{X}_2) = \boldsymbol{\mu}_2$

The covariance matrix

$$\underbrace{\begin{bmatrix} \mathbf{I}_m & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}}_{\boldsymbol{\Sigma}} \underbrace{\begin{bmatrix} \mathbf{I}_m & \mathbf{0}' \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{I}_p \end{bmatrix}}_{\mathbf{A}'}$$

$$= \begin{bmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

The multivariate normal distribution (cont.)

By Theorem 2.2, the random vectors $\mathbf{W}$ and $\mathbf{X}_2$ are independent

- The conditional of $\mathbf{W}_2|\mathbf{X}_2$ equals the marginal of $\mathbf{W}$

$$\mathbf{W}|\mathbf{X}_2 \sim N_m\big(\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\big)$$

Because of this independence,

$$\begin{aligned} &\mathbf{W} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{X}_2|\mathbf{X}_2 \\ &\qquad \sim N_m\big(\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{X}_2, \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\big) \end{aligned}$$

■

The multivariate normal distribution (cont.)

Example

Suppose that $(X, Y) \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_2^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

- $\rightsquigarrow$ μ_1 and σ_1^2 are the mean and variance of Y $(= X_1)$
- $\rightsquigarrow$ μ_2 and σ_2^2 are the mean and variance of X $(= X_2)$

$$N_m\big(\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{X}_2, \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\big)$$

The expression shows the conditional distribution of Y given $X = x$

$$N\Big[\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1), \sigma_2^2(1 - \rho^2)\Big]$$

The multivariate normal distribution (cont.)

The coefficient of x in the conditional mean of $E(Y|x)$ is $\rho\sigma_2/\sigma_1$

$$E(Y|x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}$$

The conditional mean of Y, given that $X = x$ is linear in x

The σ_1 and σ_2 are the respective standard deviations

ρ is the correlation coefficient of X and Y

Remark

This follows from the fact that the coefficient of x in a linear conditional mean $E(Y|x)$ is the product of correlation coefficient and σ_2/σ_1

The multivariate normal distribution (cont.)

The mean of the conditional distribution of Y given $X = x$

- It depends upon x (unless $\rho = 0$)

The variance $\sigma_2^2(1 - \rho^2)$ is the same for all real values of x

Thus, given that $X = x$, the conditional probability that Y is within $(\approx 2.57)\sigma_2\sqrt{(1 - \rho^2)}$ units of the conditional mean is 0.99, whatever is x

Most of the probability for the distribution of (X, Y) lies within the band

$$\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1) \pm (\approx 2.57)\sigma_2\sqrt{1 - \rho^2}$$

about the plot of the linear conditional mean

■

The multivariate normal distribution (cont.)

Theorem

Suppose $\mathbf{X}$ *has a* $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *distribution*

Let $\boldsymbol{\Sigma}$ *is positive definite*

Then, the RV $W = (\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma} (\mathbf{X} - \boldsymbol{\mu})$ *has a* $\chi^2(n)$ *distribution*

Proof

Write $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2} \boldsymbol{\Sigma}^{1/2}$, with $\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Gamma}' \boldsymbol{\Lambda}^{1/2} \boldsymbol{\Gamma}$

Then, $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2} (\mathbf{X} - \boldsymbol{\mu})$ is $N_n(\mathbf{0}, \mathbf{I}_n)$

Let $W = \mathbf{Z}'\mathbf{Z} = \sum_{i=1}^n Z_i^2$

The multivariate normal distribution (cont.)

Each Z_i has a $N(0,1)$ distribution, for $i = 1, 2, \dots, n$

From Theorem 1.1, it follows that Z_i^2 has a $\chi^2(1)$ distribution

Variables $Z_1, \dots, Z_n$ are independent standard normal RVs

- Thus, $\sum_{i=1}^n Z_i^2 = W$ has a $\chi^2(n)$ distribution
- (by an earlier corollary)

$\blacksquare$

An application

The normal distribution

Principal components analysis

An application

Principal components analysis

We consider applications of the multivariate normal distribution

- Principal components analysis (PCA)

It results in a linear function of a multivariate normal random vector

- The function preserves the 'total variation in the problem'
- The random vector has independent components

Principal components analysis (cont.)

Let the random vector $\mathbf{X}$ has the multivariate normal distribution $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- $\boldsymbol{\Sigma}$ is positive definite

Consider the spectral decomposition of $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}' \boldsymbol{\Lambda} \boldsymbol{\Gamma}$

⇝ The eigenvalues form the main diagonal of $\boldsymbol{\Lambda}$

$$\lambda_1, \lambda_2, \dots, \lambda_n$$

⇝ The corresponding eigenvectors are the columns of $\boldsymbol{\Gamma}$

$$\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$$

Assume without loss of generality that the eigenvalues are sorted

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$$

Principal components analysis (cont.)

Define the random vector $\mathbf{Y} = \mathbf{\Gamma}(\mathbf{X} - \boldsymbol{\mu})$

$\mathbf{\Gamma}\mathbf{\Sigma}\mathbf{\Gamma}' = \mathbf{\Lambda}$, thus $\mathbf{Y}$ has a $N_n(\mathbf{0}, \mathbf{\Lambda})$ distribution (by Theorem 2.1)

- The components $Y_1, Y_2, \dots, Y_n$ are independent RVs
- Y_i has a $N(0, \lambda_i)$ distribution, for $i = 1, 2, \dots, n$

The random vector $\mathbf{Y}$ is the vector of **principal components**

Principal components analysis (cont.)

Total variation of a random vector sums the variances of its components

For a random vector $\mathbf{X}$, because $\mathbf{\Gamma}$ is an orthogonal matrix

$$TV(\mathbf{X}) = \sum_{i=1}^{n} \sigma_i^2 = \text{Tr}(\mathbf{\Sigma}) = \text{Tr}(\mathbf{\Gamma}'\mathbf{\Lambda}\mathbf{\Gamma}) = \text{Tr}(\mathbf{\Lambda}\mathbf{\Gamma}\mathbf{\Gamma}') = \sum_{i=1}^{n} \lambda_i = TV(\mathbf{Y})$$

Hence, $\mathbf{X}$ and $\mathbf{Y}$ share the same total variation

Principal components analysis (cont.)

Consider the first component of $\mathbf{Y}$

$$Y_1 = \mathbf{v}_1'(\mathbf{X} - \boldsymbol{\mu})$$

- This is a linear combination of the components of $\mathbf{X} - \boldsymbol{\mu}$
- Because of orthogonality, $||\mathbf{v}_1||^2 = \sum_{j=1}^{n} v_{1j}^2 = 1$

Principal components analysis (cont.)

Consider any other linear combination of $(\mathbf{X} - \boldsymbol{\mu})$

$$\mathbf{a}'(\mathbf{X} - \boldsymbol{\mu}), \text{ with } ||\mathbf{a}||^2 = 1$$

As $\mathbf{a} \in \mathcal{R}^n$ and because $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ form a basis for $\mathcal{R}^n$

- We must have

$$\mathbf{a} = \sum_{j=1}^{n} a_j \mathbf{v}_j$$

  for some set of scalars $a_1, a_2, \dots, a_n$

Furthermore, the basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is orthonormal

$$\mathbf{a}'\mathbf{v}_i = \Big(\sum_{j=1}^{n} a_j \mathbf{v}_j\Big)'\mathbf{v}_i = \sum_{j=1}^{n} a_j \mathbf{v}_j'\mathbf{v}_i = a_i$$

Principal components analysis (cont.)

We have that $\mathbf{\Sigma} = \mathbf{\Gamma}'\mathbf{\Lambda}\mathbf{\Gamma} = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i'$ and that $\lambda_i > 0$

Then, the inequality

$$\begin{aligned} \text{Var}(\mathbf{a}'\mathbf{X}) &= \mathbf{a}'\text{Cov}(\mathbf{X})\mathbf{a} = \mathbf{a}'\mathbf{\Sigma}\mathbf{a} = \sum_{i=1}^{n} \lambda_i (\mathbf{a}'\mathbf{v}_i)^2 \\ &= \sum_{i=1}^{n} \lambda_i a_i^2 \leq \lambda_1 \sum_{i=1}^{n} a_i^2 = \lambda_1 = \text{Var}(Y_1) \end{aligned} \tag{31}$$

Y_1 has the maximum variance of any other linear combination

$\rightsquigarrow$ Y_1 is called the **first principal component** of $\mathbf{X}$

Principal components analysis (cont.)

Other components $Y_2, Y_3, \dots, Y_n$ share a similar property

Relative to the order of their associated eigenvalue

$\rightsquigarrow$ **Second**, **third**, ..., **n-th principal component**

Principal components analysis (cont.)

Theorem

For $j = 2, \dots, n$ *and* $i = 1, 2, \dots, j-1$,

$$Var[\mathbf{a}'\mathbf{X}] \leq \lambda_j = Var(Y_j)$$

for all vectors $\mathbf{a}$ *such that* $\mathbf{a} \perp \mathbf{v}_i$ *and* $||\mathbf{a}|| = 1$

Proof

The proof follows the lines of that for the first principal component ($\star$)

■