Multivariate Normal Distribution

Max Turgeon

STAT 7200–Multivariate Statistics

Building the multivariate density i

- Let $Z \sim N(0,1)$ be a standard (univariate) normal random variable. Recall that its density is given by

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

- Now if we take $Z_1, \ldots, Z_p \sim N(0,1)$ independently distributed, their joint density is

Building the multivariate density ii

$$
\begin{aligned}
\phi(z_1, \ldots, z_p) &= \prod_{i=1}^p \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z_i^2\right) \\
&= \frac{1}{(\sqrt{2\pi})^p} \exp\left(-\frac{1}{2}\sum_{i=1}^p z_i^2\right) \\
&= \frac{1}{(\sqrt{2\pi})^p} \exp\left(-\frac{1}{2}\mathbf{z}^T\mathbf{z}\right),
\end{aligned}
$$

where $\mathbf{z} = (z_1, \ldots, z_p)$.

- More generally, let $\mu \in \mathbb{R}^p$ and let Σ be a $p \times p$ positive definite matrix.

Building the multivariate density iii

- Let $\Sigma = LL^T$ be the Cholesky decomposition for Σ.

- Let $\mathbf{Z} = (Z_1, \ldots, Z_p)$ be a standard (multivariate) normal random vector, and define $\mathbf{Y} = L\mathbf{Z} + \mu$. We know from a previous lecture that
  - $E(\mathbf{Y}) = LE(\mathbf{Z}) + \mu = \mu$;
  - $\mathrm{Cov}(\mathbf{Y}) = L\mathrm{Cov}(\mathbf{Z})L^T = \Sigma$.
- To get the density, we need to compute the inverse transformation:

$$\mathbf{Z} = L^{-1}(\mathbf{Y} - \mu).$$

Building the multivariate density iv

- The Jacobian matrix J for this transformation is simply L^{-1}, and therefore

$$\begin{aligned}
|\det(J)| &= |\det(L^{-1})| \\
&= \det(L)^{-1} \qquad \text{(positive diagonal elements)} \\
&= \sqrt{\det(\Sigma)}^{-1} \\
&= \det(\Sigma)^{-1/2}.
\end{aligned}$$

Building the multivariate density v

- Plugging this into the formula for the density of a transformation, we get

$$
\begin{aligned}
f(y_1, \ldots, y_p) &= \frac{1}{\det(\Sigma)^{1/2}} \phi(L^{-1}(\mathbf{y} - \mu)) \\
&= \frac{1}{\det(\Sigma)^{1/2}} \left(\frac{1}{(\sqrt{2\pi})^p} \exp\left(-\frac{1}{2} (L^{-1}(\mathbf{y} - \mu))^T L^{-1}(\mathbf{y} - \mu) \right) \right) \\
&= \frac{1}{\det(\Sigma)^{1/2} (\sqrt{2\pi})^p} \exp\left(-\frac{1}{2} (\mathbf{y} - \mu)^T (LL^T)^{-1} (\mathbf{y} - \mu) \right) \\
&= \frac{1}{\sqrt{(2\pi)^p |\Sigma|}} \exp\left(-\frac{1}{2} (\mathbf{y} - \mu)^T \Sigma^{-1} (\mathbf{y} - \mu) \right).
\end{aligned}
$$

Example i

```r
set.seed(123)

n <- 1000; p <- 2
Z <- matrix(rnorm(n*p), ncol = p)

mu <- c(1, 2)
Sigma <- matrix(c(1, 0.5, 0.5, 1), ncol = 2)
L <- t(chol(Sigma))
```

Example ii

```r
Y <- L %*% t(Z) + mu
Y <- t(Y)

colMeans(Y)
```

```
## [1] 1.016128 2.044840
```

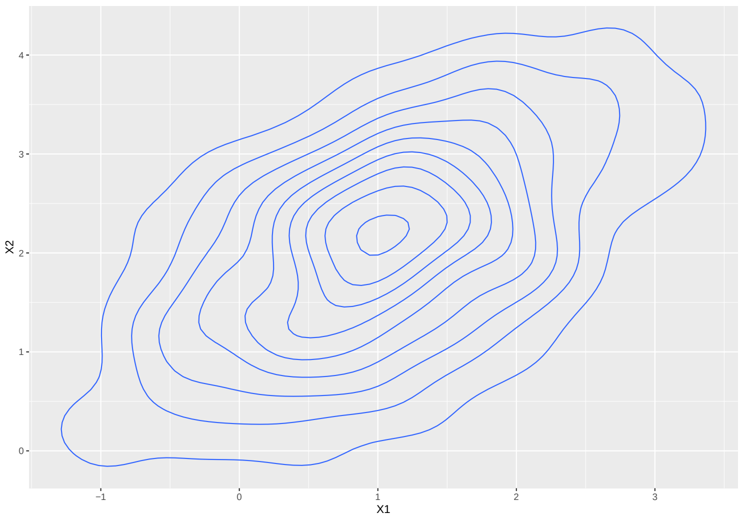
```r
cov(Y)
```

```
##           [,1]      [,2]
## [1,] 0.9834589 0.5667194
## [2,] 0.5667194 1.0854361
```

Example iii

```r
library(tidyverse)
Y %>%
  data.frame() %>%
  ggplot(aes(X1, X2)) +
  geom_density_2d()
```

Example iv



Example v

```r
library(mvtnorm)

Y <- rmvnorm(n, mean = mu, sigma = Sigma)

colMeans(Y)
```

```
## [1] 0.9812102 1.9829380
```

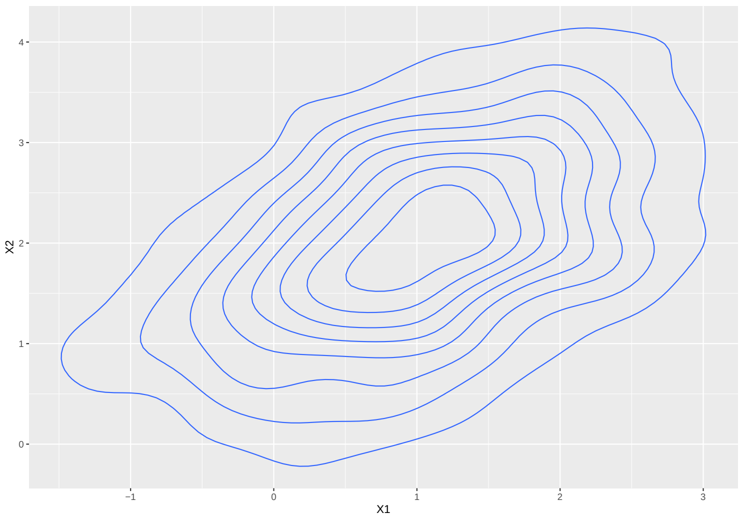
```r
cov(Y)
```

Example vi

```
##              [,1]      [,2]
## [1,] 0.9982835 0.4906990
## [2,] 0.4906990 0.9489171
```

```
Y %>%
  data.frame() %>%
  ggplot(aes(X1, X2)) +
  geom_density_2d()
```

Example vii



Characteristic function i

- Using a similar strategy, we can derive the characteristic function of the multivariate normal distribution.
- Recall that the characteristic function of the univariate standard normal distribution is given by

$$\varphi(t) = \exp\left(\frac{-t^2}{2}\right).$$

Characteristic function ii

- Therefore, if we have $Z_1, \ldots, Z_p \sim N(0,1)$ independent, the characteristic function of $\mathbf{Z} = (Z_1, \ldots, Z_p)$ is

$$\begin{aligned}
\varphi_{\mathbf{Z}}(\mathbf{t}) &= \prod_{i=1}^{p} \exp\left(\frac{-t_i^2}{2}\right) \\
&= \exp\left(\sum_{i=1}^{p} \frac{-t_i^2}{2}\right) \\
&= \exp\left(\frac{-\mathbf{t}^T\mathbf{t}}{2}\right).
\end{aligned}$$

Characteristic function iii

- For $\mu \in \mathbb{R}^p$ and $\Sigma = LL^T$ positive definite, define $\mathbf{Y} = L\mathbf{Z} + \mu$. We then have

$$\begin{aligned}
\varphi_{\mathbf{Y}}(\mathbf{t}) &= \exp\left(i\mathbf{t}^T\mu\right)\varphi_{\mathbf{Z}}(L^T\mathbf{t}) \\
&= \exp\left(i\mathbf{t}^T\mu\right)\exp\left(\frac{-(L^T\mathbf{t})^T(L^T\mathbf{t})}{2}\right) \\
&= \exp\left(i\mathbf{t}^T\mu - \frac{\mathbf{t}^T\Sigma\mathbf{t}}{2}\right).
\end{aligned}$$

Alternative characterization

A p-dimensional random vector $\mathbf{Y}$ is said to have a multivariate normal distribution if and only if every linear combination of $\mathbf{Y}$ has a *univariate* normal distribution. - **Note**: In particular, every component of $\mathbf{Y}$ is also normally distributed.

Proof i

This result follows from the Cramer-Wold theorem. Let $\mathbf{u} \in \mathbb{R}^p$. We have

$$\begin{aligned} \varphi_{\mathbf{u}^T\mathbf{Y}}(t) &= \varphi_{\mathbf{Y}}(t\mathbf{u}) \\ &= \exp\left(it\mathbf{u}^T\mu - \frac{\mathbf{u}^T\Sigma\mathbf{u}t^2}{2}\right). \end{aligned}$$

This is the characteristic function of a univariate normal variable with mean $\mathbf{u}^T\mu$ and variance $\mathbf{u}^T\Sigma\mathbf{u}$.

Proof ii

Conversely, assume $\mathbf{Y}$ has mean μ and Σ, and assume $\mathbf{u}^T\mathbf{Y}$ is normally distributed for all $\mathbf{u} \in \mathbb{R}^p$. In particular, we must have

$$\varphi_{\mathbf{u}^T\mathbf{Y}}(t) = \exp\left(it\mathbf{u}^T\mu - \frac{\mathbf{u}^T\Sigma\mathbf{u}t^2}{2}\right).$$

Now, let's look at the characteristic function of $\mathbf{Y}$:

Proof iii

$$
\begin{aligned}
\varphi_{\mathbf{Y}}(\mathbf{t}) &= E\left(\exp\left(i\mathbf{t}^T\mathbf{Y}\right)\right) \\
&= E\left(\exp\left(i(\mathbf{t}^T\mathbf{Y})\right)\right) \\
&= \varphi_{\mathbf{t}^T\mathbf{Y}}(1) \\
&= \exp\left(i\mathbf{t}^T\mu - \frac{\mathbf{t}^T\Sigma\mathbf{t}}{2}\right).
\end{aligned}
$$

This is the characteristic function we were looking for. $\square$

Counter-Example i

- Let $\mathbf{Y}$ be a mixture of two multivariate normal distributions $\mathbf{Y}_1, \mathbf{Y}_2$ with mixing probability p.
- Assume that

$$\mathbf{Y}_i \sim N_p(0, (1-\rho_i)I_p + \rho_i \mathbf{1}\mathbf{1}^{\mathbf{T}}),$$

where $\mathbf{1}$ is a p-dimensional vector of 1s.

  - In other words, the diagonal elements are 1, and the off-diagonal elements are ρ_i.

Counter-Example ii

- First, note that the characteristic function of a mixture distribution is a mixture of the characteristic functions:

$$\varphi_{\mathbf{Y}}(\mathbf{t}) = p\varphi_{\mathbf{Y}_1}(\mathbf{t}) + (1-p)\varphi_{\mathbf{Y}_2}(\mathbf{t}).$$

- Therefore, unless $p = 0, 1$ or $\rho_1 = \rho_2$, the random vector $\mathbf{Y}$ does **not** follow a normal distribution.
- But the components of a mixture are the mixture of each component.
  - Therefore, all components of $\mathbf{Y}$ are univariate standard normal variables.

Counter-Example iii

- In other words, **even if all the margins are normally distributed, the joint distribution may not follow a multivariate normal**.

Useful properties i

- If $\mathbf{Y} \sim N_p(\mu, \Sigma)$, A is a $q \times p$ matrix, and $b \in \mathbb{R}^q$, then

$$A\mathbf{Y} + b \sim N_q(A\mu + b, A\Sigma A^T).$$

- If $\mathbf{Y} \sim N_p(\mu, \Sigma)$ then all subsets of $\mathbf{Y}$ are normally distributed; that is, write
  - $\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}$, $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$;
  - $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$.
  - Then $\mathbf{Y}_1 \sim N_r(\mu_1, \Sigma_{11})$ and $\mathbf{Y}_2 \sim N_{p-r}(\mu_2, \Sigma_{22})$.

Useful properties ii

- Assume the same partition as above. Then the following are equivalent:
  - $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are independent;
  - $\Sigma_{12} = 0$;
  - $\mathrm{Cov}(\mathbf{Y}_1, \mathbf{Y}_2) = 0$.

Exercise (J&W 4.3)

Let $(Y_1, Y_2, Y_3) \sim N_3(\mu, \Sigma)$ with $\mu = (3, 1, 4)$ and

$$\Sigma = \begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Which of the following random variables are independent? Explain.

1. Y_1 and Y_2.
2. Y_2 and Y_3.
3. (Y_1, Y_2) and Y_3.
4. $0.5(Y_1 + Y_2)$ and Y_3.
5. Y_2 and $Y_2 - \frac{5}{2}Y_1 - Y_3$.

Conditional Normal Distributions i

- **Theorem**: Let $\mathbf{Y} \sim N_p(\mu, \Sigma)$, where
  - $\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}$, $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$;
  - $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$.
- Then the *conditional distribution* of $\mathbf{Y}_1$ given $\mathbf{Y}_2 = \mathbf{y}_2$ is multivariate normal $N_r(\mu_{1|2}, \Sigma_{1|2})$, where
  - $\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2)$
  - $\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Proof i

Let B be the same dimension as Σ_{12}. We have

$$
\begin{aligned}
&\begin{pmatrix} I & -B \\ 0 & I \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} I & 0 \\ -B^T & I \end{pmatrix} \\
=& \begin{pmatrix} \Sigma_{11} - B\Sigma_{21} - \Sigma_{12}B^T + B\Sigma_{22}B^T & \Sigma_{12} - B\Sigma_{22} \\ \Sigma_{21} - \Sigma_{22}B^T & \Sigma_{22} \end{pmatrix}.
\end{aligned}
$$

Proof ii

If we take $B = \Sigma_{12}\Sigma_{22}^{-1}$, the two off-diagonal blocks become 0, which implies that the blocks are independent. Also, the left-upper block simplifies to

$$\Sigma_{1|2} = \Sigma_{11} + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

On the other hand, we also have

Proof iii

$$
\begin{pmatrix} I & -B \\ 0 & I \end{pmatrix} \mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{Y}_2 \\ \mathbf{Y}_2 \end{pmatrix}
$$
$$
\begin{pmatrix} I & -B \\ 0 & I \end{pmatrix} \mu = \begin{pmatrix} \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2 \\ \mu_2 \end{pmatrix}.
$$

We can therefore conclude

$$
\begin{aligned}
\mathbf{Y}_1 - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{Y}_2 = \mathbf{Y}_1 - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{y}_2 \mid \mathbf{Y}_2 = \mathbf{y}_2 \\
\sim N(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2, \Sigma_{1|2}).
\end{aligned}
$$

Proof iv

By adding $\Sigma_{12}\Sigma_{22}^{-1}\mathbf{y}_2$ to the left-hand side, the result follows. $\square$

Conditional Normal Distributions ii

- **Corrolary**: Let $\mathbf{Y}_2 \sim N_{p-r}(\mu_2, \Sigma_{22})$ and assume that $\mathbf{Y}_1$ given $\mathbf{Y}_2 = y_2$ is multivariate normal $N_r(Ay_2 + b, \Omega)$, where Ω does not depend on y_2. Then $\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} \sim N_p(\mu, \Sigma)$, where
    - $\mu = \begin{pmatrix} A\mu_2 + b \\ \mu_2 \end{pmatrix}$;
    - $\Sigma = \begin{pmatrix} \Omega + A\Sigma_{22}A^T & A\Sigma_{22} \\ \Sigma_{22}A^T & \Sigma_{22} \end{pmatrix}$.

Exercise

- Let $\mathbf{Y}_2 \sim N_1(0,1)$ and assume

$$\mathbf{Y}_1 \mid \mathbf{Y}_2 = y_2 \sim N_2\left(\begin{pmatrix} y_2 + 1 \\ 2y_2 \end{pmatrix}, I_2\right).$$

  Find the joint distribution of $(\mathbf{Y}_1, \mathbf{Y}_2)$.

Another important result i

- Let $\mathbf{Y} \sim N_p(\mu, \Sigma)$, and let $\Sigma = LL^T$ be the Cholesky decomposition of Σ.
- We know that $\mathbf{Z} = L^{-1}(\mathbf{Y} - \mu)$ is normally distributed, with mean 0 and covariance matrix

$$\mathrm{Cov}(\mathbf{Z}) = L^{-1}\Sigma(L^{-1})^T = I_p.$$

- Therefore $(\mathbf{Y} - \mu)^T\Sigma^{-1}(\mathbf{Y} - \mu)$ is the sum of *squared* standard normal random variables.
  - In other words, $(\mathbf{Y} - \mu)^T\Sigma^{-1}(\mathbf{Y} - \mu) \sim \chi^2(p)$.
  - This can be seen as a generalization of the univariate result $\left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi^2(1)$.

Another important result ii

- From this, we get a result about the probability that a multivariate normal falls within an *ellipse*:

$$P\left((\mathbf{Y}-\mu)^T\Sigma^{-1}(\mathbf{Y}-\mu)\leq\chi^2(\alpha;p)\right)=1-\alpha.$$

  - We can use this to construct a confidence region around the sample mean.