Multivariate normal distribution

Surajit Ray
Reader, University of Glasgow
Univariate normal distribution

Univariate normal with mean 2 and variance 1
Density shape of a bivariate normal
Bivariate normal density - 3D density plot

\[ \mu = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 2 \end{pmatrix} \]
Bivariate normal density - contour plot

$$\mu = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 2 \end{pmatrix}$$
Bivariate normal density with a different mean

\[ \mu = \begin{pmatrix} -1 \\ -3 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 2 \end{pmatrix} \]
Bivariate normal density with a different variance

$$\mu = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$
Bivariate normal density with strong correlation

\[ \mu = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0.95 \\ 0.95 & 1 \end{pmatrix} \]
## Functions for statistical distributions in R

<table>
<thead>
<tr>
<th>Normal</th>
<th>Multivariate Normal</th>
<th>t</th>
<th>Multivariate t</th>
</tr>
</thead>
<tbody>
<tr>
<td>rnorm</td>
<td>rmvnorm</td>
<td>rt</td>
<td>rmvt</td>
</tr>
<tr>
<td>dnorm</td>
<td>dmvnorm</td>
<td>dt</td>
<td>dmvt</td>
</tr>
<tr>
<td>pnorm</td>
<td>pmvnorm</td>
<td>pt</td>
<td>pmvt</td>
</tr>
<tr>
<td>qnorm</td>
<td>qmvnorm</td>
<td>qt</td>
<td>qmvvt</td>
</tr>
</tbody>
</table>
Functions for statistical distributions in R

- The first letter denotes
  - `p` for "probability"
  - `q` for "quantile"
  - `d` for "density"
  - `r` for "random"

- Followed by the distribution name
  - `norm`
  - `mvnorm`
  - `t`
  - `mvt`
The `rmvnorm` function

```r
library(mvtnorm)
rmvnorm(n, mean, sigma)
```

Need to specify:

- `n` the number of samples
- `mean` the mean of the distribution
- `sigma` the variance-covariance matrix
Using rmvnorm to generate random samples

Generate 1000 samples from a 3 dimensional normal with

\[ \mu = \begin{pmatrix} 1 \\ 2 \\ -5 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 5 \end{pmatrix} \]

```r
mul <- c(1, 2, -5)
sigma <- matrix(c(1,1,0,
                 1,2,0,
                 0,0,5),3,3)
set.seed(34)
rmvnorm(n = 1000, mean = mul, sigma = sigma)
```
Plot of generated samples
Let's practice simulating from a multivariate normal distribution!
Density of a multivariate normal distribution

Surajit Ray
Reader, University of Glasgow
Why calculate the density of a distribution?
Why calculate the density of a distribution?
Univariate normal functions `dnorm()`
Probability density of a bivariate normal

Standard bivariate normal with

\[ \mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \]

Density heights calculated at several locations (xy coordinates)
Density using `dmvnorm`

```r
library(mvtnorm)
dmvnorm(x, mean, sigma)
```

- `x` can be a row vector or a matrix

```r
mu1 <- c(1, 2)
sigma1 <- matrix(c(1, .5, .5, 2), 2)
dmvnorm(x = c(0, 0), mean = mu1, sigma = sigma1)
```

0.0384
Density at multiple points using `dmvnorm`

```r
x <- rbind(c(0, 0), c(1, 1), c(0, 1)); x

```
```
[1,] 0 0
[2,] 1 1
[3,] 0 1
```

```r
dmvnorm(x = x, mean = mu, sigma = sigma)
```
```
[1] 0.0384 0.0904 0.0679
```
Plotting bivariate densities with perspective plot

Steps:

- Create grid of $x$ and $y$ coordinates
- Calculate density on grid
Plotting bivariate densities with perspective plot

Steps:

- Create grid of $x$ and $y$ coordinates
- Calculate density on grid
- Convert densities into a matrix
- Create perspective plot using `persp()` function
Code for plotting bivariate densities

```r
# Create grid
d <- expand.grid(seq(-3, 6, length.out = 50), seq(-3, 6, length.out = 50))

# Calculate density on grid
dens1 <- dmvnorm(as.matrix(d), mean=c(1,2), sigma=matrix(c(1, .5, .5, 2), 2))

# Convert to matrix
dens1 <- matrix(dens1, nrow = 50)

# Use perspective plot
persp(dens1, theta = 80, phi = 30, expand = 0.6, shade = 0.2,
      col = "lightblue", xlab = "x", ylab = "y", zlab = "dens")
```
Changing viewing angle in perspective plot

persp() with theta = 30, phi = 30  

persp() with theta = 80, phi = 10
Let's practice!
Cumulative Distribution and Inverse CDF
When do we need to calculate CDF and inverse CDF?
When do we need to calculate CDF and inverse CDF?

Normal density with $\mu = 210$ and $\sigma = 10$
When do we need to calculate CDF and inverse CDF?

Area under the curve for $x < 200$
When do we need to calculate CDF and inverse CDF?

\[ \text{pnorm}(200, \text{mean} = 210, \text{sd} = 10) \]

[1] 0.159
When do we need to calculate CDF and inverse CDF?

What is the \( x_0 \) such that the cumulative probability at \( x_0 \) is 0.95?

\[
\text{qnorm}(\ p = 0.95, \ \text{mean} = 210, \ \text{sd} = 10) \\
[1] \ 226.45
\]

\[ \Rightarrow \text{95\% of the coffee jars will have less than 226.45 grams of coffee} \]
Cumulative distribution for a bivariate normal

Bivariate CDF at $x = 2$ and $y = 4$ for a normal with $\mu = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $\Sigma = \begin{pmatrix} 1 & .5 \\ .5 & 2 \end{pmatrix}$
Cumulative distribution using pmvnorm

Bivariate CDF at x = 2 and y = 4 for a normal with \( \mu = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \), \( \Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 2 \end{pmatrix} \)

```r
mul <- c(1, 2)
sigma <- matrix(c(1, 0.5, 0.5, 2), 2)

pmvnorm(upper = c(2, 4), mean = mul, sigma = sigma)
```

```
[1] 0.79
attr("error")
[1] 1e-15
attr("msg")
[1] "Normal Completion"
```
Probability between two values using pmvnorm

Probability of $1 < x < 2$ and $2 < y < 4$

```r
pmvnorm(lower = c(1, 2),
        upper = c(2, 4),
        mean = mu1,
        sigma = sigma1)
```
Probability between two values using `pmvnorm`

Probability of $1 < x < 2$ and $2 < y < 4$

```
pmvnorm(lower = c(1, 2),
        upper = c(2, 4),
        mean = mu1,
        sigma = sigmal)
```

[1] 0.163
Inverse CDF for bivariate normal

Dark red ellipse is the 0.95 quantile
Implementing `qmvnorm` to calculate quantiles

```r
# Create a diagonal covariance matrix
sigma <- diag(2)
sigma

[,1] [,2]
[1,] 1 0
[2,] 0 1

# Calculate the quantile
qmvnorm(p = 0.95, sigma = sigma, tail = "both")

$quantile
[1] 2.24

$f.quantile
[1] -1.31e-06

attr("message")
[1] "Normal Completion"
```

The red circle with radius 2.24 contains 0.95 of the probability.
Let's practice!
Checking normality of multivariate data

Surajit Ray
Reader, University of Glasgow
Why check normality?

- Classical statistical techniques that assume univariate/multivariate normality:
  - Multivariate regression
  - Discriminant analysis
  - Model-based clustering
  - Principal component analysis (PCA)
  - Multivariate analysis of variance (MANOVA)
Review: univariate normality tests

```
qqnorm(iris_raw[, 1])
qqline(iris_raw[, 1])
```

- If the values lie along the reference line the distribution is close to normal
Review: univariate normality tests

```r
qqnorm(iris_raw[, 1])
qqline(iris_raw[, 1])
```

- If the values lie along the reference line the distribution is close to normal
- Deviation from the line might indicate
  - heavier tails
  - skewness
  - outliers
  - clustered data
qqnorm of all variables

uniPlot(iris_raw[, 1:4])
MVN library multivariate normality test functions

- Multivariate normality tests by
  - Mardia
  - Henze-Zirkler
  - Royston

- Graphical approaches
  - chi-square Q-Q
  - perspective
  - contour plots
MVN library multivariate normality test functions

- Multivariate normality tests by
  - Mardia ✓
  - Henze-Zirkler ✓
  - Royston

- Graphical approaches
  - chi-square Q-Q ✓
  - perspective
  - contour plots
Using `mardiaTest` to check multivariate normality

```r
mardiaTest(iris_raw[,1:4])
```

Mardia Multivariate Normality Test
----------------------------------
data : iris_raw[, 1:4]

  glp : 2.697
chi.skew : 67.43
p.value.skew : 4.758e-07

  g2p : 23.74
z.kurtosis : -0.2301
p.value.kurt : 0.818

  chi.small.skew : 69.33
p.value.small : 2.342e-07

Result : Data are not multivariate normal.
Using \texttt{qqplot} from \texttt{mardiaTest} to check multivariate normality

\texttt{mardiaTest(iris\_raw[, 1:4], qqplot = TRUE)}
Using `hzTest` to check multivariate normality

```r
hzTest(iris_raw[,1:4])
```

<table>
<thead>
<tr>
<th>Henze-Zirkler's Multivariate Normality Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>data : iris_raw[, 1:4]</td>
</tr>
<tr>
<td>HZ   : 2.333269</td>
</tr>
<tr>
<td>p-value : 0</td>
</tr>
<tr>
<td>Result : Data are not multivariate normal.</td>
</tr>
</tbody>
</table>

---
Testing multivariate normality by species

```r
mardiaTest(iris[iris_raw$Species == "setosa", 1:4])
```

**Mardia's Multivariate Normality Test**

<table>
<thead>
<tr>
<th>g1p</th>
<th>3.08</th>
</tr>
</thead>
<tbody>
<tr>
<td>chi.skew</td>
<td>25.7</td>
</tr>
<tr>
<td>p.value.skew</td>
<td>0.177</td>
</tr>
<tr>
<td>g2p</td>
<td>26.5</td>
</tr>
<tr>
<td>z.kurtosis</td>
<td>1.29</td>
</tr>
<tr>
<td>p.value.kurt</td>
<td>0.195</td>
</tr>
<tr>
<td>chi.small.skew</td>
<td>27.85973</td>
</tr>
<tr>
<td>p.value.small</td>
<td>0.1127617</td>
</tr>
</tbody>
</table>

**Result**: Data are multivariate normal.
Let's make use of the tests for multivariate normality!