Properties of the multivariate normal

We can write that a vector is multivariate normal as $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
Some important properties of multivariate normal distributions include

1. Linear combinations of the variables $y_1, \ldots, y_p$ are also normal with
Note that for some distributions, such as the Poisson, sums of independent (but not necessarily identically distributed) random variables stay within the same family of distributions. For other distributions, they don't stay in the same family (e.g., exponential random variables). However, it is not clear that sums of two correlated Poissons will still be Poisson. Also, differences between Poisson random variables are not Poisson. For the normal, we have the nice property that even if two (or more) normal random variables are correlated, any linear combinations will still be normal.

Properties of multivariate normal distributions

2. If $\mathbf{A}$ is constant (entries are not random variables) and is $q \times p$ with rank $q \leq p$, then

$$\mathbf{Ay} \sim N_q(\mathbf{A}\mu, \mathbf{A\Sigma A}')$$

What happens if $q > p$?

Properties of multivariate normal distributions

3. A vector $\mathbf{y}$ can be standardized using either

$$\mathbf{z} = (\mathbf{T}')^{-1}(\mathbf{y} - \boldsymbol{\mu})$$

where $\mathbf{T}$ is obtained using the Cholesky decomposition so that $\mathbf{T}'\mathbf{T} = \boldsymbol{\Sigma}$, or

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

This standardization is similar to the idea of z-scores; however, just taking the usual z-scores of the individual variables in $\mathbf{y}$ will still leave the variables correlated. The standardizations above result in

$$z \sim N_p(\mathbf{0}, \mathbf{I})$$

Properties of multivariate normal distributions

4. Sums of squares of p independent standard normal random variables have a χ^2 distribution with p degrees of freedom. Therefore, if $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

$$\begin{aligned}
(\mathbf{y}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu}) &= (\mathbf{y}-\boldsymbol{\mu})'\left(\Sigma^{1/2}\Sigma^{1/2}\right)^{-1}(\mathbf{y}-\boldsymbol{\mu}) \\
&= (\mathbf{y}-\boldsymbol{\mu})'\left(\Sigma^{1/2}\right)^{-1}\left(\Sigma^{1/2}\right)^{-1}(\mathbf{y}-\boldsymbol{\mu}) \\
&= \left[\left(\Sigma^{1/2}\right)^{-1}(\mathbf{y}-\boldsymbol{\mu})\right]'\left(\Sigma^{1/2}\right)^{-1}(\mathbf{y}-\boldsymbol{\mu}) \\
&= \mathbf{z}'\mathbf{z}
\end{aligned}$$

Here the vector $\mathbf{z}$ consists of i.i.d. standard normal vectors according to property 3, so $\mathbf{z}'\mathbf{z}$ is a sum of squared i.i.d. standard normals, which is known to have χ^2 distribution. Therefore

$$(\mathbf{y}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu}) \sim \chi^2_p$$

Properties of multivariate normal distributions

5. Normality of marginal distributions

If $\mathbf{y}$ has p random variables and is multivariate normal, then any subset $y_{i_1}, \ldots, y_{i_r}$, $r < p$, is also multivariate normal. We can assume that the r variables of interested are listed first so that

$$\mathbf{y}_1 = (y_1, \ldots, y_r)', \quad \mathbf{y}_2 = (y_{r+1}, \ldots, y_p)'$$

Then we have

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

and

$$\mathbf{y}_1 \sim N_r(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$$

Properties of multivariate normal distributions

Note that if y_1 and y_2 are each normal, it doesn't follow that $\mathbf{y} = (y_1, y_2)'$ is multivariate normal. For an extreme example, let $y_1 \sim N(0, 1)$, $z \sim \text{Bernoulli}(1/2)$ and $y_2 = y_1 \cdot I(z = 1) - y_1 \cdot (z = 0)$. In other words, with probability $1/2$, $y_2 = y_1$, and with probability $1/2$, $y_2 = -y_1$. Then y_2 is normal, yet $(y_1, y_2)'$ is not multivariate normal.

What does the distribution of $(y_1, y_2)'$ look like?

Properties of multivariate normal distributions

```
> y2 <- y1*(z==1)-y1*(z==0)
> y1 <- rnorm(1000)
> z <- rbinom(1000,1,.5)
> y2 <- y1*(z==1)-y1*(z==0)
> shapiro.test(y2) # quick test of normality of a vector

Shapiro-Wilk normality test

data:  y2
W = 0.9989, p-value = 0.8234
```

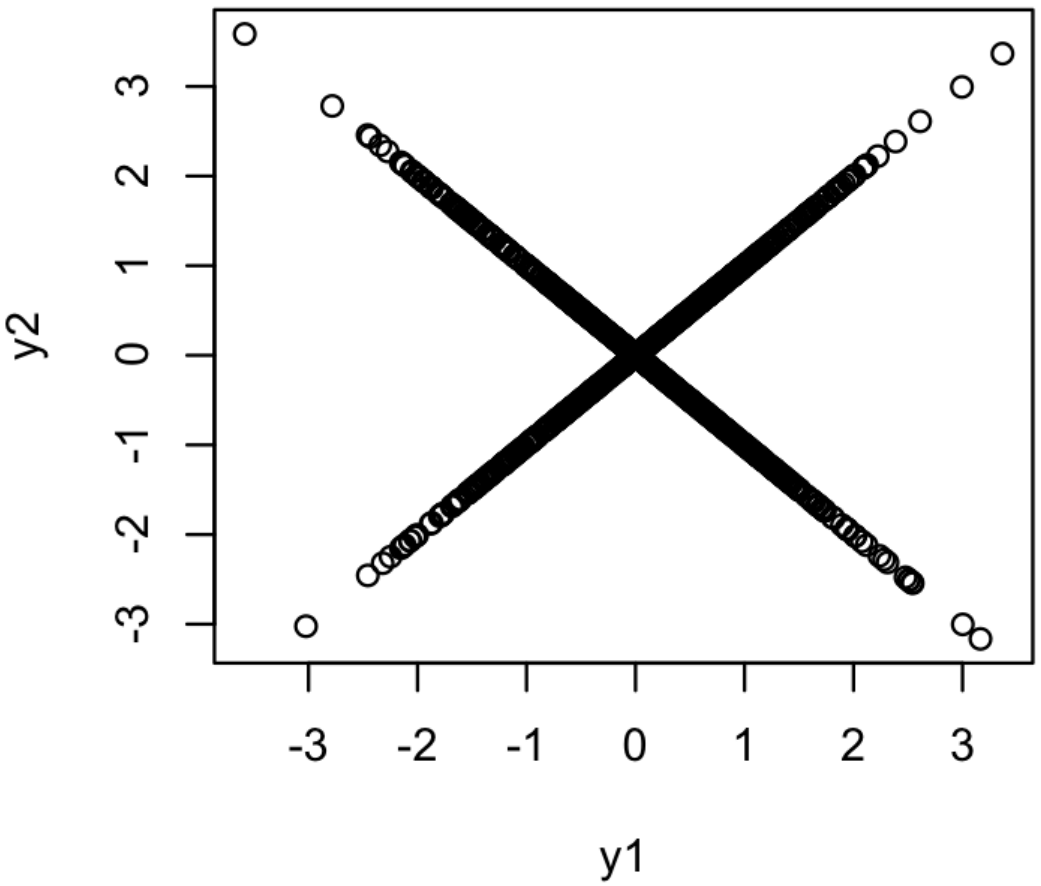
Properties of multivariate normal distributions

To check more theoretically that y_2 is normal, we use the fact that for a standard normal y_1 and $-y_1$ have the same distribution

$$\begin{aligned} P(y_2 \leq x) &= P(y_2 \leq x|z=1)P(z=1) + P(y_2 \leq x|z=2)P(z=2) \\ &= P(y_1 \leq x)(1/2) + P(-y_1 \leq x)(1/2) \\ &= P(y_1 \leq x)(1/2) + P(y_1 \leq x)(1/2) \\ &= P(y_1 \leq x) \end{aligned}$$

Therefore y_2 has the same CDF (cumulative distribution function) as y_1, so they have the same distribution. This shows more than that y_2 is normal — it also standard normal just like y_1.

Properties of multivariate normal distributions



Marginal distributions

If you have a collection of random variables, and you ignore some of them, the distribution of the remaining is a marginal distribution. For a bivariate random variable $\mathbf{y} = (y_1, y_2)'$, the distribution of y_1 is a marginal distribution of the distribution of $\mathbf{y}$.

In non-vector notation, the joint density for two random variables is often written

$$f_{12}(y_1, y_2)$$

and the marginal distribution can be obtained by

$$f_1(y_1) = \int_{-\infty}^{\infty} f_{12}(y_1, y_2)\, dy_2$$

The joint density for $\mathbf{y}_1$ is

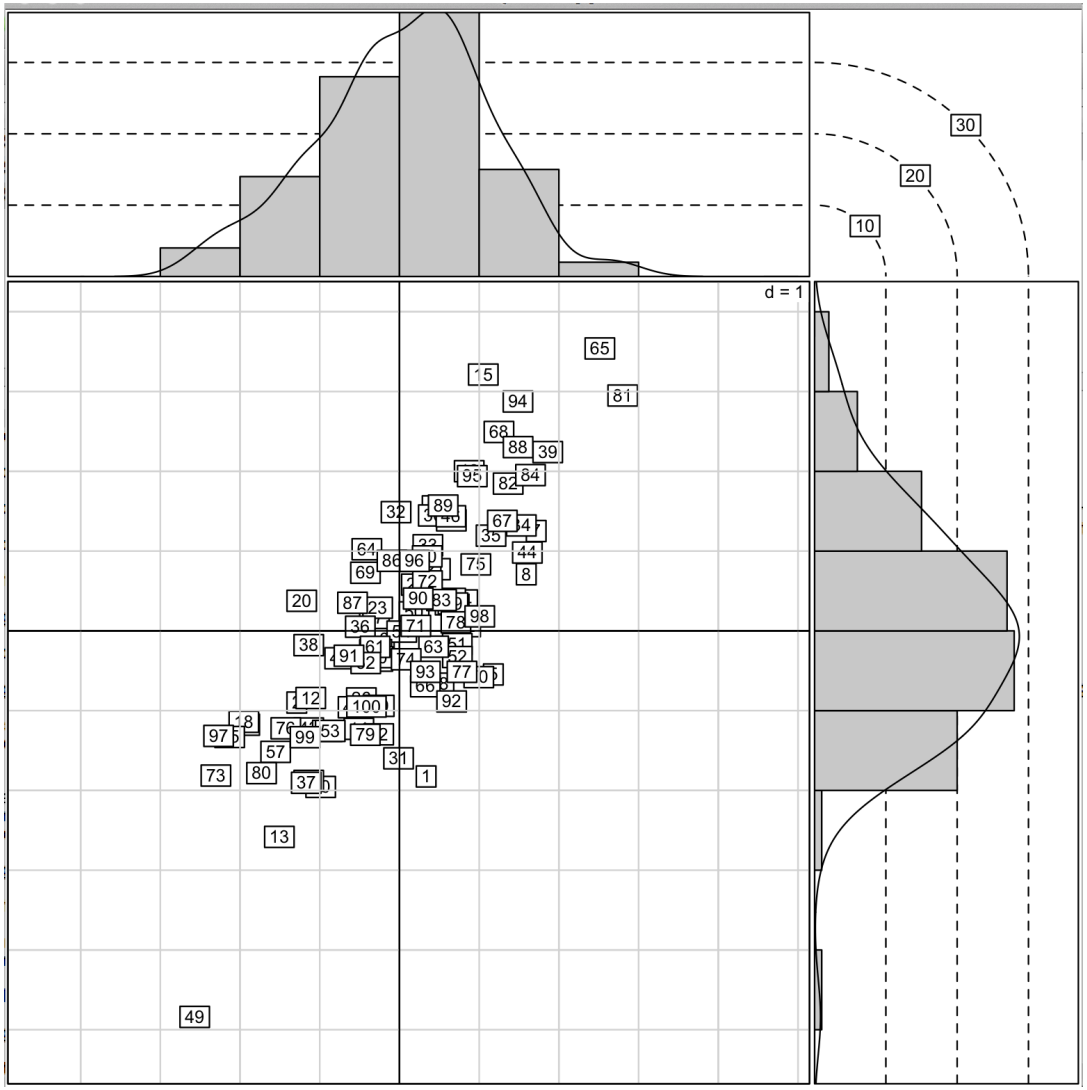
$$f_{12\cdots r}(y_1, \ldots, y_r) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{1\cdots p}(y_1, \ldots, y_p)\, dy_{r+1} \cdots dy_p$$

And this is why it is called a marginal density.

Plotting marginal densities in R

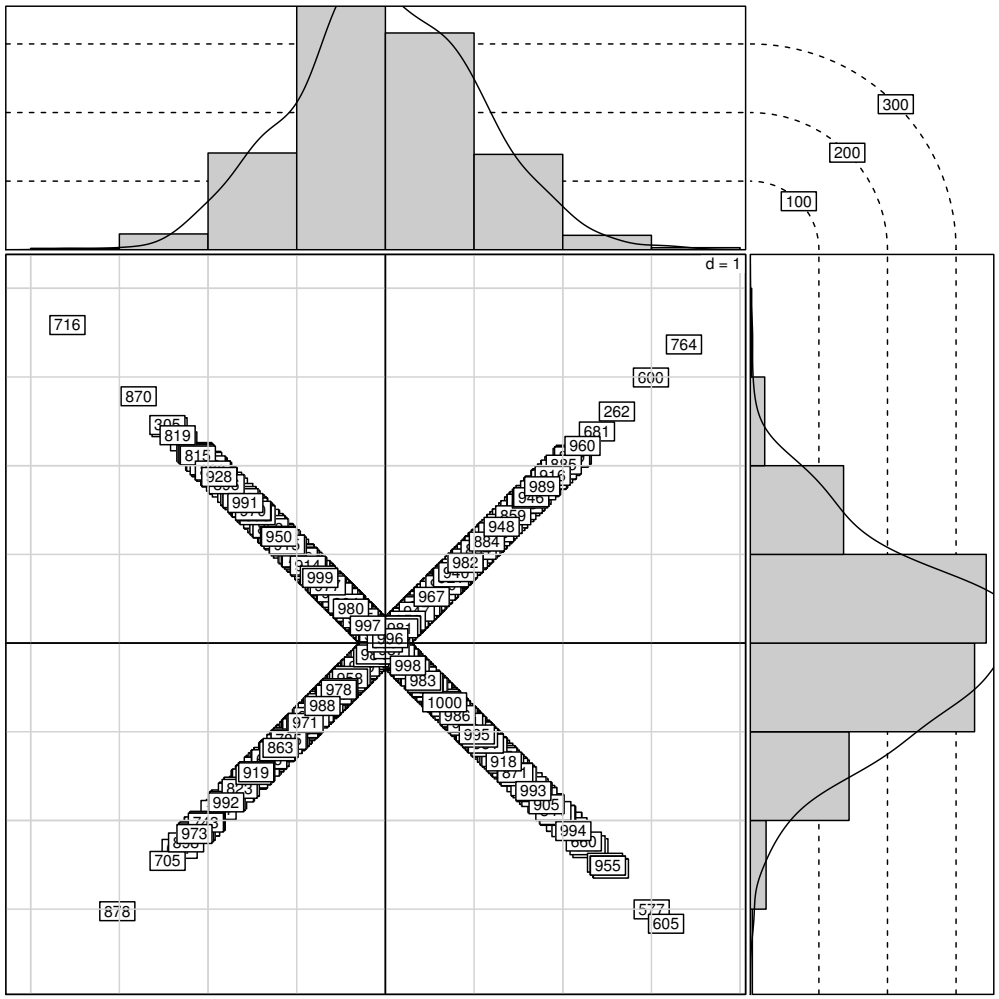
```
> install.packages("ade4")
> library(ade4)
> x <- rnorm(100)
> y <- x+ rnorm(100)
> d <- data.frame(x,y)
> s.hist(d)
```

Plotting marginal densities in R



Plotting marginal densities in R

For the weird example with $y_2 = y_1 1 \cdot I(z = 1) - y_1 \cdot I(z = 0)$



Properties of the multivariate normal distribution

Let the observation vector (rows of the data matrix) be partitioned into $\mathbf{y}$ and $\mathbf{x}$ with

$$E\begin{pmatrix}\mathbf{y}\\ \mathbf{x}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\mu}_y\\ \boldsymbol{\mu}_x\end{pmatrix}, \quad cov\begin{pmatrix}\mathbf{y}\\ \mathbf{x}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx}\\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{\mathbf{xx}}\end{pmatrix}$$

with

$$\begin{pmatrix}\mathbf{y}\\ \mathbf{x}\end{pmatrix} \sim N_{p+q}\left[\begin{pmatrix}\boldsymbol{\mu}_y\\ \boldsymbol{\mu}_x\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx}\\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{\mathbf{xx}}\end{pmatrix}\right]$$

Properties of the multivariate normal distribution

5. (a) The subvectors $\mathbf{y}$ and $\mathbf{x}$ are independent if $\mathbf{\Sigma}_{yx} = \mathbf{0}$.

(b) y_i and y_j (or y_i and x_j) are independent if $\sigma_{ij} = 0$.
These properties do not always hold if the distribution is not multivariate normal. In particular, for the weird example where
$y_2 = y_1 \cdot I(z = 1) - y_1 \cdot I(z = 0)$, what can you say about the correlation between y_1 and y_2, and what can you say about their independence?

Properties of the multivariate normal distribution

It is often easier to show that two variables are uncorrelated than that they are independent. So this property of the multivariate normal, that no correlation implies independence, is quite useful.

Conditional distributions

Given two or more random variables with a joint distribution, we can condition on some random variables to get the conditional distribution of the remaining variables. For example, with the heights and ages of couples example, we could look at the distribution of couples heights given that the husband is six feet tall, or the distribution of the heights given that both partners are exactly 50.

For continuous random variables, note that we can condition on events that have probability 0.

Conditional distributions

The conditional density is often notated using, for example $f(y|x)$. Conditional and marginal densities are related by

$$f(y|x) = f(y,x)/f(x)$$

In other words, the conditional density is the joint density divided by a marginal density. This idea can be extended to more than two random variables, for example

$$f(y_1, y_2|y_3, y_4, y_5) = \frac{f(y_1, y_2, y_3, y_4, y_5)}{f(y_3, y_4, y_5)}$$

The notation here is a little sloppy since I am using f to denote several different densities (functions). We might write this instead as

$$f_{12|345}(y_1, y_2|y_3, y_4, y_5) = \frac{f_{12345}(y_1, y_2, y_3, y_4, y_5)}{f_{345}(y_3, y_4, y_5)}$$

Properties of the multivariate normal distribution

6. If $\mathbf{y}$ and $\mathbf{x}$ are not independent, then $cov(\mathbf{y}, \mathbf{x}) = \mathbf{\Sigma}_{yx} \neq \mathbf{0}$ and the conditional density $f(\mathbf{y}|\mathbf{x})$ is multivariate normal (or normal if $\mathbf{y}$ has one component) with

$$E(\mathbf{y}|\mathbf{x}) = \boldsymbol{\mu}_y + \mathbf{\Sigma}_{yx}\mathbf{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$$

$$cov(\mathbf{y}|\mathbf{x}) = \mathbf{\Sigma}_{yy} - \mathbf{\Sigma}_{yx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xy}$$

7. If $\mathbf{y}$ and $\mathbf{x}$ are independent and the same size (e.g., $p \times 1$), then

$$\mathbf{y} + \mathbf{x} = N_p(\boldsymbol{\mu}_y + \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{yy} + \boldsymbol{\Sigma}_{xx})$$

$$\mathbf{y} - \mathbf{x} = N_p(\boldsymbol{\mu}_y - \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{yy} + \boldsymbol{\Sigma}_{xx})$$

Estimation for the multivariate normal

Just as we estimate parameters of the normal distribution, using $\overline{y}$ for μ and s^2 for σ^2, so we will want to estimate parameters of the multivariate normal. The usual sample statistics of sample means, sample variances and covariances can also be used to estimate the parameters of the multivariate normal distribution.

A common way to estimate parameters in statistics is called **maximum likelihood**. The idea is to find the paramter values that maximize the joint density of the data. The maximum likelihood estimators are close to the usual sample statistics except that for the variance for a single normal random variable, the maximum likelihood estimator is

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2$$

where we usually use $\frac{1}{n-1}$ instead of $\frac{1}{n}$. This is because the maximum likelihood estimator is biased (it tends to underestimate the variance), and using $n-1$ makes the estimate unbiased.

Estimation for the multivariate normal

Similarly, the maximum likelihood estimate for the covariances will use the usual sample estimate but with $\frac{1}{n}$ in place of $\frac{1}{n-1}$.

Estimation for the multivariate normal

For the univariate case, we can consider the distribution of sample statistics. For example, if $y_1, \ldots, y_n$ are iid Normal(μ, σ^2), then

$$\overline{y} \sim N(\mu, \sigma^2/n)$$

The sample variance will also tend to vary for different samples, and it too has a sampling distribution. The sampling distribution of s^2 is described by

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

That is, scaling the sample variance in a certain way results in a χ^2 distribution where the degrees of freedom depend on the sample size. This is related to the fact that a sum of squared standard normal variables has a χ^2 distribution.

Estimation for the multivariate normal

We'd like to extend these results to the multivariate normal.

If each of n independent observations comes from the same p-dimension multivariate normal distribution, then we have $\mathbf{y}_1, \ldots, \mathbf{y}_n \sim N_p(\boldsymbol{\mu}, \Sigma)$, so $\bar{\mathbf{y}} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$.

From the Multivariate Central Limit Theorem, we have that if $\mathbf{y}_1, \ldots, \mathbf{y}_n$ are nonnormal but are iid, then for typical distrbutions (where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are defined), the result still holds approximately for large samples (large n). We can also write that

Estimation for the multivariate normal

The sample covariance matrix $\mathbf{S}$ also has a sampling distribution. The scaled version $(n-1)\mathbf{S}$ has what is called a Wishart distribution, which gives a distribution on square matrices. This is denoted by

$$(n-1)\mathbf{S} \sim W_p(n-1, \mathbf{\Sigma})$$

where $n-1$ denotes the degrees of freedom and p is the size of the covariance matrix. You can think of the Wishart distribution as the multivariate analogue of the χ^2 distribution.
We have

$$(\mathbf{y} - \bar{\mathbf{y}})(\mathbf{y} - \bar{\mathbf{y}})' \sim W_p(n-1, \mathbf{\Sigma})$$

$$(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})' \sim W_p(n, \mathbf{\Sigma})$$

so that estimating the mean reduces the degrees of freedom by 1

Estimation for the multivariate normal

In the univariate case, when sampling from a normal distribution

$$\overline{y} \text{ and } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$$

are independent (even though $\overline{y}$ appears in both). The extends to the multivariate case where

$$\overline{\mathbf{y}} \text{ and } \mathbf{S}$$

are independent.

Testing for normality

To use techniques that assume normality of data, it is good to be able to test for normality, although techniques might be applied anyway even when such tests fail. In particular, it often isn't reasonable to think that the data is really normally distributed, but if departures from normality aren't too severe, than many methods will work well anyway.

For these tests, the null hypothesis is that the data is normal (or multivariate normal), and the alternative hypothesis is that the data does not come from a normal distribution.

First we'll go over some univariate tests of normality.

Testing for normality

A standard graphical approach for the univariate case are Q-Q plots, which plot the ordered observations against their expected values. If the data is

$$y_1, y_2, \ldots, y_n$$

then we reorder the data as

$$y_{(1)}, y_{(2)}, \ldots, y_{(n)}$$

which are called **order statistics**. For example, $y_{(1)}$ is the minimum, $y_{(2)}$ is the second smallest observation (or is tied for smallest), and $y_{(n)}$ is the maximum observed value. If n is odd, then $y_{(n/2)}$ is the sample median. The order statistics $y_{(i)}$ each have a distribution, which can be determined theoretically. We then make a scatterplot of y_i against $E(y_{(i)})$ for each i. If the data is approximately normal, then the scatterplot should fall roughly on a line.

Testing for normality

$Q - Q$ plots for normal data can be generated in R using the `qqnorm()` function.

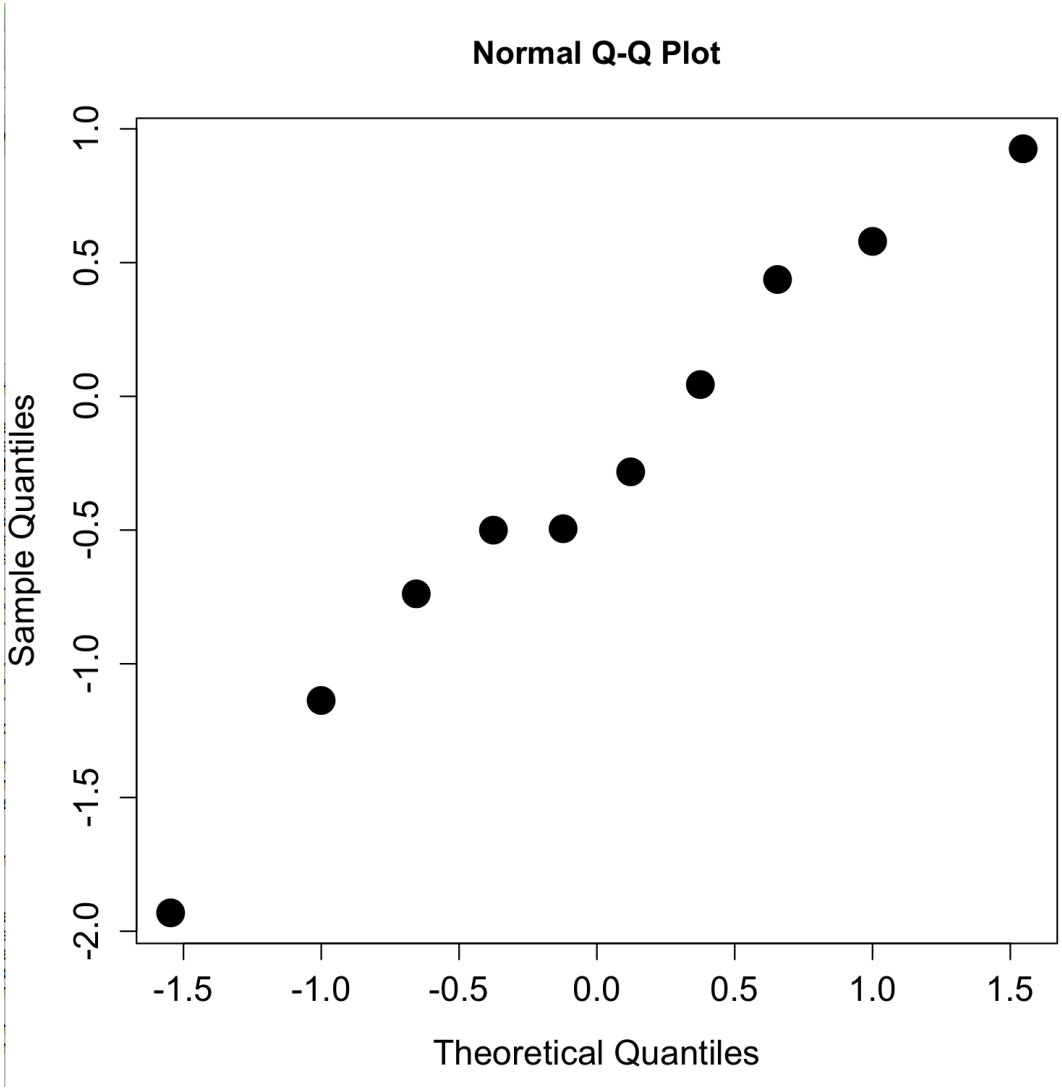
```
> x <- rnorm(10)
> a <- qqnorm(x,cex=2.4,pch=16)
> a
$x
 [1] -1.5466353 -0.6554235  0.1225808  0.6554235  1.0004905 -0.3754
 [7]  0.3754618 -0.1225808  1.5466353 -1.0004905

$y
 [1] -1.93129894 -0.73845624 -0.28232207  0.43687829  0.57934180 -0
 [7]  0.04397096 -0.49506508  0.92562595 -1.13723152

> cor(a$x,a$y)
[1] 0.98716
```

It isn't necessary to save the results of `qqnorm(y)` to an object, but doing so allows you to get extra information

Q-Q plot in R


Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5

Testing for normality

The correlation is close to 1.0 if the observed order statistics are close to the expected order statistics. For different samples, the correlation will be slightly different, so the correlation has some sampling distribution. The Shapiro-Wilk test uses the distribution of the squared correlation to test whether the correlation is low enough to have evidence of non-normality.

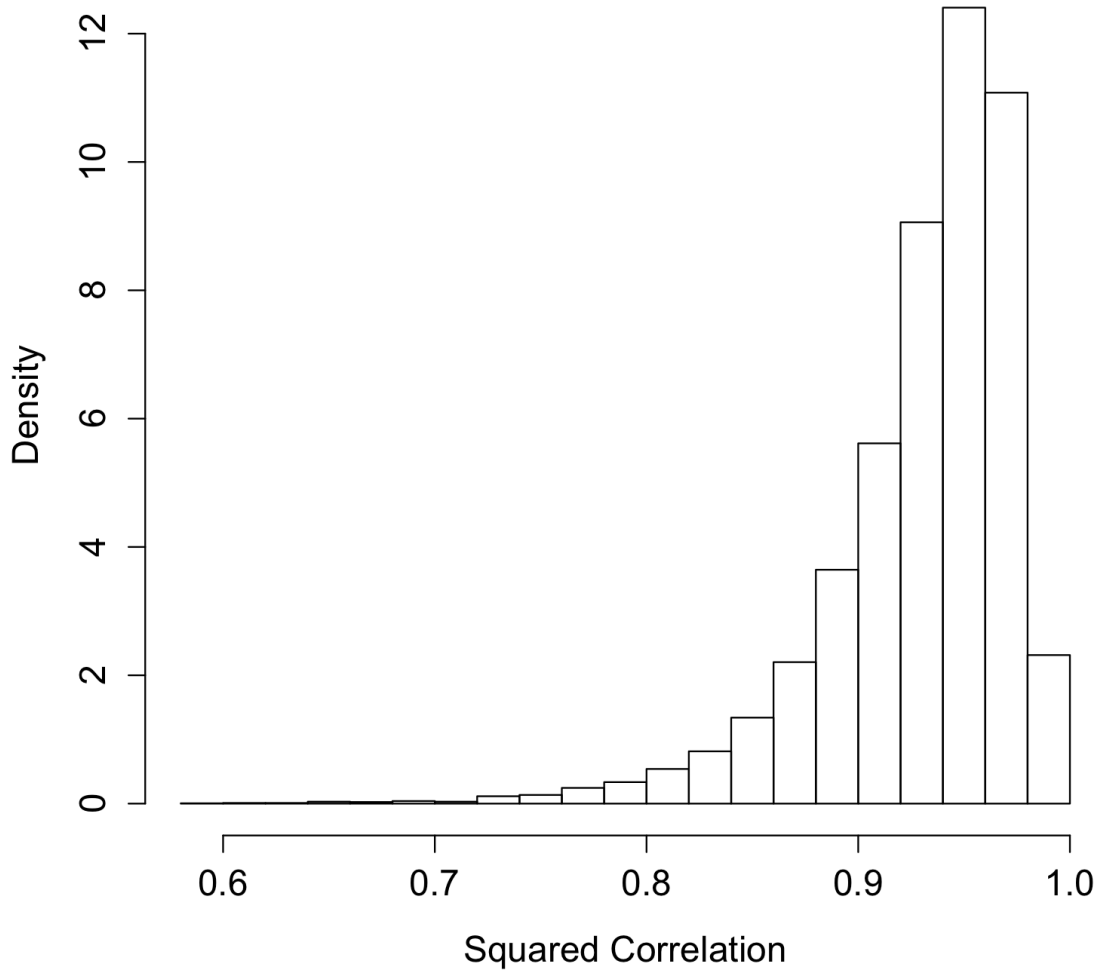
Testing for normality

Repeating random samples of size 10 gives you an idea of the distribution of the correlation.

```
> x <- rnorm(10)
> a <- qqnorm(x)
> cor(a$x,a$y)
[1] 0.9842459
> x <- rnorm(10)
> a <- qqnorm(x)
> cor(a$x,a$y)
[1] 0.9136584
> x <- rnorm(10)
> a <- qqnorm(x)
> cor(a$x,a$y)
[1] 0.9667297
```

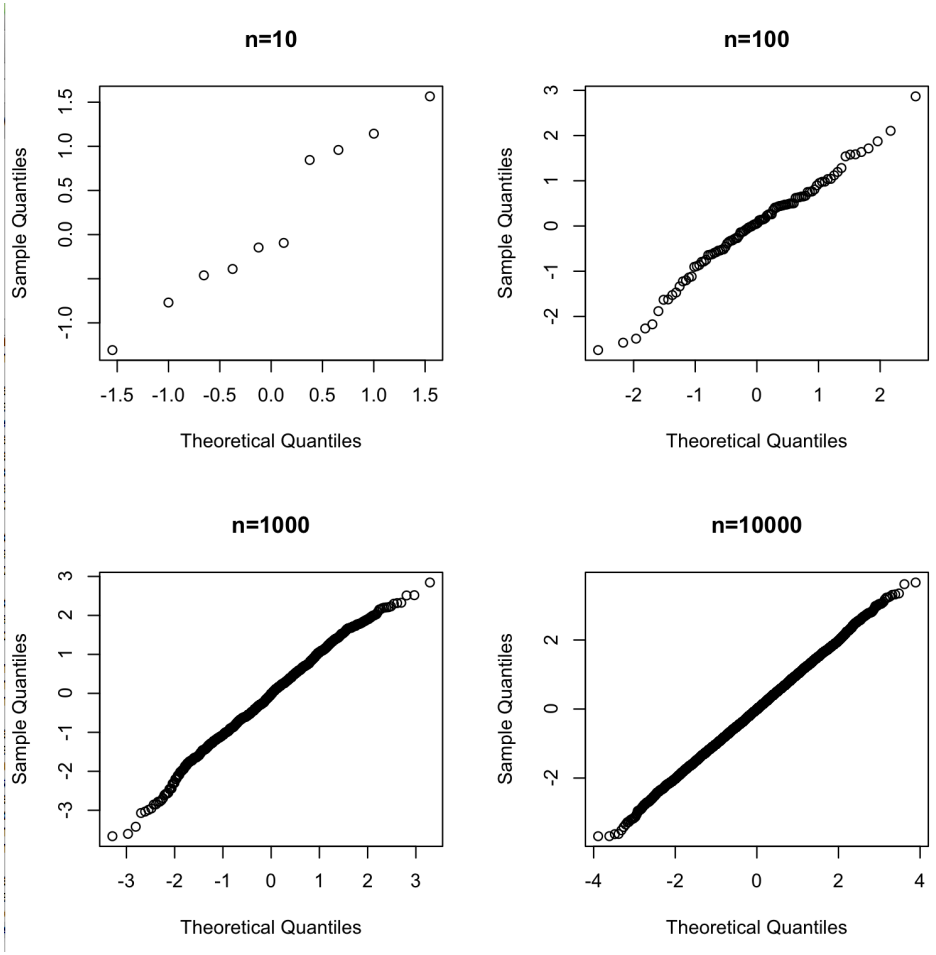
Distribution of the Shapiro-Wilk statistic

You reject if the test statistic is sufficiently small.



Q-Q plots

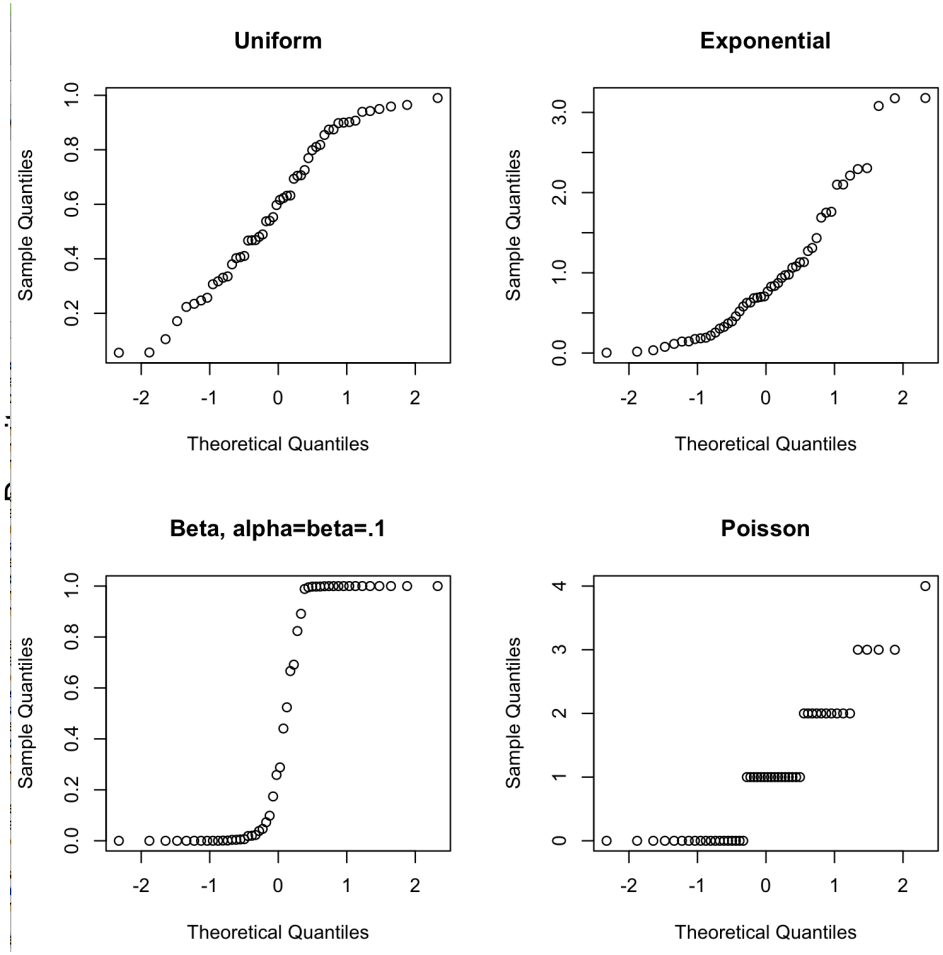
The distribution of the Shapiro-Wilks statistic (and the correlation in the Q-Q plot) depends on the sample size as well as the distribution of the data.



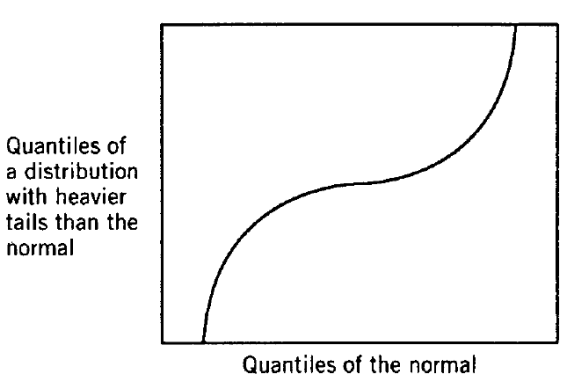
Q-Q plots

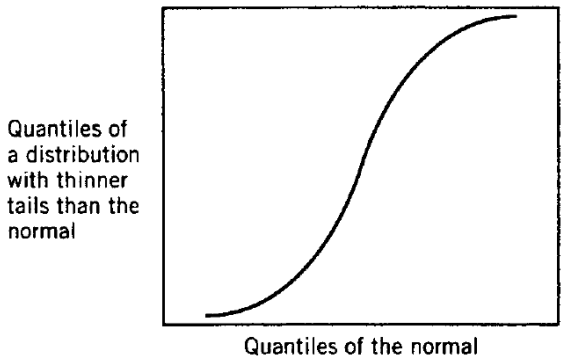
The idea of Q-Q plots makes sense for other distributions as well, but is most commonly used to check for normality. To test whether data is exponentially distributed (for a particular exponential distribution), you could plot the ordered data against the expected order statistics and simulate the distribution of the correlation (or squared correlation) to check whether your data is consistent with the theoretical values.

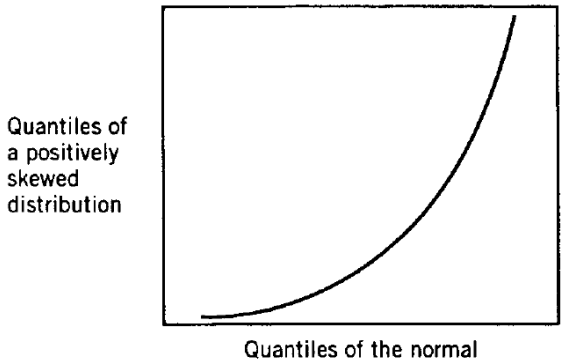
Q-Q plots for non-normal data



Q-*Q* plots for non-normal data



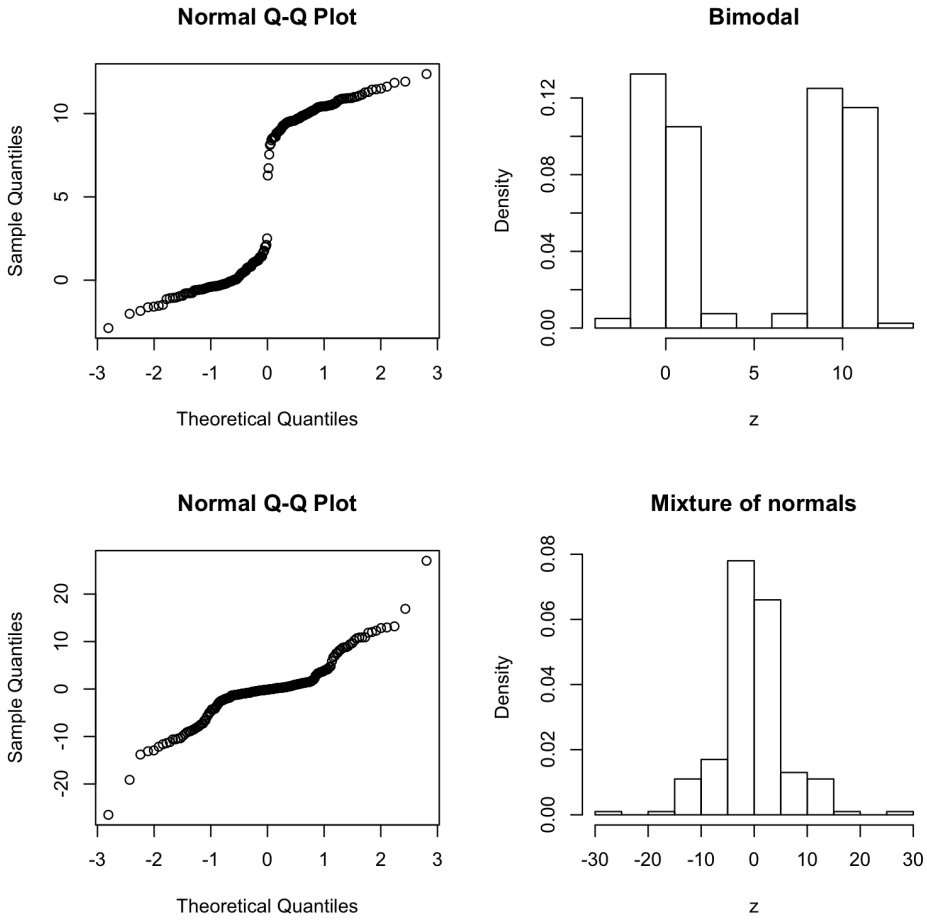




Other examples of non-normality

```
> x <- rnorm(100)
> y <- rnorm(100,10)
> z <- c(x,y)
> qqnorm(z)
> hist(z,main="Bimodal",prob=T)
> x <- rnorm(100)
> y <- rnorm(100,0,10)
> z <- c(x,y)
> qqnorm(z)
> hist(z,main="Mixture of normals",prob=T)
```

Q-Q plots for non-normal data



Other tests of normality

You can also test for normality based on skewness and kurtosis, the third and fourth moments of a distribution:

$$\sqrt{b_1} = \frac{\sqrt{n}\sum_{i=1}^{n}(y_i - \overline{y})^3}{[\sum_{i=1}^{n}(y_i - \overline{y})^2]^{3/2}}$$

$$b_2 = \frac{n\sum_{i=1}^{n}(y_i - \overline{y})^4}{[\sum_{i=1}^{n}(y_i - \overline{y})^2]^2}$$

Just as $\overline{y}$ and s have sampling distributions, so do b_1 and b_2. Comparing the observed value of b_1 to its expected value tests whether the observed skewness in the distribution is significantly different from expected, and comparing the observed value of b_2 to its expected value tests whether a distribution is more (or less) peaked than expected based on a normal distribution.

Other tests of normality

The parameters estimated by $\sqrt{b_1}$ and b_2 are $E[(y-\overline{y})^3]$ and $E[(y-\overline{y})^4]$, the third and fourth moments central of the distirbution. Note that SAS output regularly gives sample skewness and kurtosis from your sample using the PROC UNIVARIATE procedure.

There are tables in the appendices of the book that give critical values for $\sqrt{b_1}$ and b_2 to carry out formal tests for normality. For our purposes, using `shapiro.test()` in R is sufficient.

Other tests of normality

A goodness-of-fit test can be performed using the χ^2 distribution. The idea is to categorize the data into intervals. For the normal, you might use intervals $(-\infty, -2), [-2, -1), [-1, 0), [0, 1), [1, 2), [2, \infty)$. You can make a count of how many observations fall in each category. Then you can use probabilities of each category to determine the expected count, and use the chi-square statistic:

$$\sum_i \frac{(O_i - E_i)^2}{E_i}$$

to determine whether the observed number of observations is consistent with expectations. Here you would use a χ^2_{n-1} ($n-1$ degrees of freedom). The number of categories is somewhat arbitrary, but too many (with few observations per cell) or too few could result in loss of power (i.e., low probability of rejecting the null hypothesis when it is false).

Testing for multivariate normality

A necessary condition for multivariate normality is that the individual variables are normal; however, this is not a sufficient condition. Individual variables can be normal but not jointly multivariate normal.

Methods for testing for multivariate normality are not as well developed, however. Part of the problem is the dimensionality of the data.

A goodness-of-fit test could be done again. For bivariate data, you could partition $\mathbb{R}^2$ into rectangles and determine the expected number of observations in each rectangle and then perform a χ^2 test. However, there will be many more rectangles than intervals for one-dimension, so you might have a lot of empty cells or you might need to have coarse rectangles. Either way, this type of test can result in low power – i.e., be unlikely to reject the null hypothesis when it is false.

Testing for multivariate normality

An informal, graphical approach is to notice is to look at scatterplot matrices. The idea is that if a vector is multivariate, then so are its subsets. In particular, subsets of size 2 are bivariate normal. If any of the plots in a scatterplot appear not to be bivariate normal due to outliers, nonlinear trends, bimodality, etc., then multivariate normality is violated.

Apparently bivariate normality for all subsets doesn't gaurantee multivariate normality, but it is a more thorough check than only considering univariate normality.

Testing for multivariate normality

One approach for testing for multivariate normality is to define

$$D_i = (\mathbf{y}_i - \bar{\mathbf{y}})' \mathbf{S}^{-1} (\mathbf{y}_i - \bar{\mathbf{y}})$$

(the squared Mahalonibis distance of each observation vector from the mean observation).

If the data are multivariate normal, then

$$u_i = \frac{n D_i^2}{n-1}$$

has a beta distribution, and a Q-Q plot can be made comparing the ordered values $u_{(1)}, \ldots, u_{(n)}$ to their expected values.

Slightly simpler is to use $D^2_{(n)}$ (the maximum squared Mahalanobis distance), and compare to a critical value. The critical values are given in Table A.6 for $p \leq 5$ dimensions.

Testing for multivariate normality

Another approach is to generalize tests based on skewness and kurtosis. Define

$$\beta_{1,p} = E[(\mathbf{y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})]^3$$

$$\beta_{1,p} = E[(\mathbf{y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})]^4$$

Then for a $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ random variable

$$\beta_{1,p} = 0, \quad \beta_{2,p} = p(p+2)$$

Testing for multivariate normality

Let

$$\widehat{\mathbf{\Sigma}} = \sum_{i=1}^{n} (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})'$$

$$g_{ij} = (\mathbf{y} - \bar{\mathbf{y}})' \widehat{\mathbf{\Sigma}}^{-1} (\mathbf{y} - \bar{\mathbf{y}})$$

$$b_{1,p} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij}^3$$

$$b_{2,p} = \frac{1}{n} \sum_{i=1}^{n} g_{ii}^2$$

Here $\widehat{\mathbf{\Sigma}}$ is the maximum likelihood estimate of $\mathbf{\Sigma}$.

From these statistics, Table A.5 can be used to test the hypothesis of multivariate normality. The book gives tables for the distribution of $b_{1,p}$ and $b_{2,p}$ which is kind of crude, for example it gives the 90th, 92.5th, 95th, 97.5th, and 99th percentiles, but not values in between, and uses sample sizes of $n = 10, 20, \ldots, 100, 150, 200, 300, 400, 600$. For values in between, you would have to interpolate the critical value.
Alternatively, you could simulate the distribution of these statistics. Using tables of critical values is rather old-fashioned and dates from times when computing power was limited or expensive (such as having to pay for time on a mainframe computer). You can always simulate values instead of relying on a table.

Simulating the multivariate normal

The multivariate normal can be simulated in R by using the `mvrnorm()` function from the `MASS` library. Here you specify the number of observations, the mean vector μ and the covariance matrix $\mathbf{\Sigma}$. Unlike `rnorm`, there are no default parameters for the mean and covariance.

Simulating the multivariate normal

```
> library(MASS)
> mu <- c(2,3)
> sigma <- matrix(c(1,2,3,4),ncol=2)
> mu
[1] 2 3
> sigma
     [,1] [,2]
[1,]    1    3
[2,]    2    4
> y <- mvrnorm(3,mu,sigma)
> y
         [,1]     [,2]
[1,] 1.241342 1.482685
[2,] 1.233697 1.467393
[3,] 1.622789 2.245578
>
```