Homework ...

Simone Gervasoni

Contents

Exercise 1

For $i = 1, \ldots, n$, let Y_i be i.i.d. random variables taking values in $\{1, 2, \ldots, p, p+1\}$ with probabilities $\pi_1, \ldots, \pi_p, \pi_{p+1} > 0$, $\sum_{j=1}^{p+1} \pi_j = 1$. If we code Y_i via the one-hot vector $Y_i = (Y_{i1}, \ldots, Y_{i,p+1})$ where $Y_{ij} = 1$ when $Y_i = j$, then $\sum_{i=1}^{n} Y_i$ has multinomial distribution, Multinomial$(n, \pi_1, \ldots, \pi_{p+1})$, and the multivariate Central Limit Theorem (CLT) implies

$$\sqrt{n}(\hat{\pi} - \pi) \xrightarrow{d} \mathcal{N}_{p+1}(0, \mathrm{diag}(\pi) - \pi\pi^T)$$

for $\hat{\pi} = n^{-1} \sum_{i=1}^{n} Y_i$ and $\pi = (\pi_1, \ldots, \pi_{p+1})$. Assume n is sufficiently large so that $\sqrt{n}(\hat{\pi} - \pi)$ is normally distributed according to the CLT above and let X be the vector of the first p coordinates.

Point 1

What is the distribution of X? Justify your answer.

Let $Y_i \in \{1, 2, \ldots, p\}$ $i = 1, \ldots, n$ with probability $\pi_1, \ldots, \pi_p, \pi_{p+1} > 0$ and $\sum_{j=1}^{p+1} \pi_j = 1$; i.e.

$$Y_i = \begin{cases} 1 & \text{with probability } \pi_1 \\ 2 & \text{with probability } \pi_2 \\ & \vdots \\ p+1 & \text{with probability } \pi_{p+1} \end{cases}$$

The one-hot vector $Y_i = (Y_{i1}, \ldots, Y_{ip+1})$ is such that $Y_i = 1$ when $Y_i = j$. So Y_i codified as one hot vector is distributed as $Y_i \sim$ Multinomial$(1, \pi_1, \ldots, \pi_{p+1})$. Further $\sum_{i=1}^{n} Y_i$ has multinational distribution $\sum_{i=1}^{n} Y_i \sim$ Multinomial$(n, \pi_1, \ldots, \pi_{p+1})$.

By the CLT we can say

$$(X_1, \ldots, X_{p+1}) \sim \mathcal{N}_{p+1}(0, \mathrm{diag}(\pi) - \pi\pi^T)$$

. Where the covariance matrix is:

$$\Omega = \begin{pmatrix} \pi_1 & & & \\ & \pi_2 & & \\ & & \ddots & \\ & & & \pi_{p+1} \end{pmatrix} - \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_{p+1} \end{pmatrix} \begin{pmatrix} \pi_1 & \ldots & \pi_{p+1} \end{pmatrix}$$

$$\begin{pmatrix} \pi_1 & & & \\ & \pi_2 & & \\ & & \ddots & \\ & & & \pi_{p+1} \end{pmatrix} - \begin{pmatrix} \pi_1^2 & \pi_1\pi_2 & \ldots & \pi_1\pi_{p+1} \\ \pi_2\pi_1 & \pi_2^2 & \ldots & \pi_2\pi_{p+1} \\ \vdots & & \ddots & \\ \pi_{p+1}\pi_1 & & & \pi_{p+1}^2 \end{pmatrix} =$$

$$\Omega = \left(\begin{array}{ccc|c} \pi_1(1-\pi_1) & -\pi_1\pi_2 & \ldots & -\pi_1\pi_{p+1} \\ -\pi_2\pi_1 & \pi_2(1-\pi_2) & \ldots & -\pi_2\pi_{p+1} \\ \vdots & & & \\ \hline -\pi_{p+1}\pi_1 & -\pi_{p+1}\pi_2 & \ldots & \pi_{p+1}(1-\pi_{p+1}) \end{array}\right)$$

If we take the first p components of a Gaussian vector of dimension p+1 , it is still a Gaussian vectorr of dimension p in fact:

$$(X_1, \ldots, X_{p+1}) \sim \mathcal{N}_{p+1}(0, \Omega)$$

* the new mean vector $\underline{0}_p$ is the vector of dimension $p < p+1$. * The element $\Omega_{p+1,p+1} = \underbrace{\pi_{p+1}}_{>0} \underbrace{(1-\pi_{p+1})}_{>^{(1)}\ 0} > 0$

(1) $1 - \pi_{p+1} > 0 \iff \pi_{p+1} < 1$ and this is always true because π_{p+1} is a probability that is defined $(0, 1)$.

So we can say thatthe marginal distribution of the Gaussian vector of dimension p+1 is still a Gaussian:

$$(X_1, \dots, X_p) \sim \mathcal{N}_p(0, \Sigma)$$

where

$$\Sigma := \Omega_{p\times p} = \begin{pmatrix} \pi_1(1-\pi_1) & -\pi_1\pi_2 & \dots & -\pi_1\pi_p \\ -\pi_2\pi_1 & \pi_2(1-\pi_2) & \dots & -\pi_2\pi_p \\ \vdots & & \ddots & \\ -\pi_p\pi_1 & -\pi_p\pi_2 & \dots & \pi_p(1-\pi_p) \end{pmatrix}$$

Point 2

Let Σ be the $p \times p$ covariance matrix of X. Find the inverse of Σ.

We want to find the inverse of the covariance matrix:

$$\Sigma = \begin{pmatrix} \pi_1^2 & -\pi_1\pi_2 & \dots & -\pi_1\pi_p \\ -\pi_2\pi_1 & \pi_2^2 & \dots & -\pi_2\pi_p \\ \vdots & \vdots & \ddots & \vdots \\ -\pi_p\pi_1 & -\pi_p\pi_2 & \dots & \pi_p^2 \end{pmatrix}$$

First of all we need to check if the matrix is invertible, so we need to prove that $det(\Sigma) \neq 0$.
Since evaluating the determinant is not quite easy, we find a **Lemma** that give us an equivalent condition for the invertibility.

Sherman Morrison Formula:

Suppose $A \in \mathbb{R}^{n\times n}$ invertible and $u, v \in \mathbb{R}$ are column vectors. $A + uv^T$ is invertible $\iff$ $1 + v^T Au \neq 0$. In this case:

$$(A + uv^T) = A^{-1} - \frac{A^{-1}uv^T A^{-1}}{1 + v^T A^{-1}u}.$$

In our case we have:

$$\Sigma = diag(\pi_1, \dots, \pi_p) - \pi\pi^T = A + uv^T$$

where: * $A = \text{diag}(\pi_1, \dots, \pi_p) \in \mathbb{R}^{p\times p}$ * $u = \begin{pmatrix} -\pi_1 \\ -\pi_2 \\ \vdots \\ -\pi_p \end{pmatrix}$ * $v^T = \begin{pmatrix} \pi_1 & \dots & \pi_p \end{pmatrix}$

Remark : in our case there is a minus so to be coherent with the notation of the formula (2) we take the minus inside the vector u.

Is A invertible?

$$|A| = \left| \begin{pmatrix} \pi_1 & & \\ & \ddots & \\ & & \pi_p \end{pmatrix} \right| = \pi_1\pi_2 \dots \pi_p > 0$$

so it is invertible beacuse the determinant is bigger than 0

To apply the **Lemma** we check that $1 + v^T A^{-1}u \neq 0$.

Since A is a diagonal matrix we know that the inverse is as such:

$$A^{-1} = \begin{pmatrix} \frac{1}{\pi_1} & & \\ & \ddots & \\ & & \frac{1}{\pi_p} \end{pmatrix}$$

now we calculate :

$$1+v^TA^{-1}u = 1+\begin{pmatrix}\pi_1 & \dots & \pi_p\end{pmatrix}\begin{pmatrix}\frac{1}{\pi_1} & & \\ & \ddots & \\ & & \frac{1}{\pi_p}\end{pmatrix}\begin{pmatrix}-\pi_1\\ \vdots \\ -\pi_p\end{pmatrix} =$$

$$= 1+\begin{pmatrix}1 & 1 & \dots & 1\end{pmatrix}\begin{pmatrix}-\pi_1\\ \vdots \\ -\pi_p\end{pmatrix} =$$
$$= 1-(\pi_1+\dots+\pi_p) \neq 0 \Leftrightarrow \pi_1+\dots+\pi_p \neq^{(2)} 1$$

(2) this is always true beacuse by hypothesis $\sum_{i=1}^{p+1}\pi_i = 1$ and $\pi_i > 0$ so $\sum_{i=1}^{p}\pi_i < 1$ because π_{p+1} is missing.

The hypotesis of the **Lemma** are verifed now we need to find

$$(A+uv^T)^{-1} = A^{-1} - \frac{A^{-1}uv^TA^{-1}}{\underbrace{1+v^tA^{-1}u}_{1-(\pi_1+\dots+\pi_p)=\pi_{p+1}}}$$

The numerator of the fraction is the following

$$A^{-1}uv^TA^{-1} = \underbrace{\begin{pmatrix}\frac{1}{\pi_1} & & \\ & \ddots & \\ & & \frac{1}{\pi_p}\end{pmatrix}}_{p\times p}\underbrace{\begin{pmatrix}-\pi_1\\ \vdots \\ -\pi_p\end{pmatrix}}_{p\times 1}\begin{pmatrix}\pi_1 & \dots & \pi_p\end{pmatrix}\begin{pmatrix}\frac{1}{\pi_1} & & \\ & \ddots & \\ & & \frac{1}{\pi_p}\end{pmatrix}$$
$$= \begin{pmatrix}-1\\ -1 \\ \vdots \\ -1\end{pmatrix}\begin{pmatrix}\pi_1 & \dots & \pi_p\end{pmatrix}\begin{pmatrix}\frac{1}{\pi_1} & & \\ & \ddots & \\ & & \frac{1}{\pi_p}\end{pmatrix}$$
$$= \begin{pmatrix}-\pi_1 & \dots & -\pi_p\\ -\pi_1 & \dots & -\pi_p\\ \vdots & & \vdots \\ -\pi_1 & \dots & -\pi_p\end{pmatrix}\begin{pmatrix}\frac{1}{\pi_1} & & \\ & \ddots & \\ & & \frac{1}{\pi_p}\end{pmatrix} = \underbrace{\begin{pmatrix}-1 & \dots & -1\\ -1 & \dots & -1\\ \vdots & & \vdots \\ -1 & \dots & -1\end{pmatrix}}_{p\times p}$$

$$\Sigma^{-1} = A^{-1} - \begin{pmatrix}-\frac{1}{\pi_{p+1}} & \dots & -\frac{1}{\pi_{p+1}}\\ \vdots & \ddots & \vdots \\ -\frac{1}{\pi_{p+1}} & \dots & -\frac{1}{\pi_{p+1}}\end{pmatrix}$$
$$= \begin{pmatrix}\frac{1}{\pi_1} & & \\ & \ddots & \\ & & \frac{1}{\pi_p}\end{pmatrix} + \begin{pmatrix}\frac{1}{\pi_{p+1}} & \dots & \frac{1}{\pi_{p+1}}\\ \vdots & \ddots & \vdots \\ \frac{1}{\pi_{p+1}} & \dots & \frac{1}{\pi_{p+1}}\end{pmatrix}$$
$$= \begin{pmatrix}\frac{1}{\pi_1}+\frac{1}{\pi_{p+1}} & \frac{1}{\pi_{p+1}} & \dots & \frac{1}{\pi_{p+1}}\\ \frac{1}{\pi_{p+1}} & \frac{1}{\pi_2}+\frac{1}{\pi_{p+1}} & \dots & \frac{1}{\pi_{p+1}}\\ \vdots & \vdots & & \vdots \\ \frac{1}{\pi_{p+1}} & \frac{1}{\pi_{p+1}} & \dots & \frac{1}{\pi_p}+\frac{1}{\pi_{p+1}}\end{pmatrix}$$

Point 3

Let $\pi = (\pi_0, \ldots, \pi_0, 1 - p\pi_0)$ for some $0 < \pi_0 < 1/p$. Find the eigenvalues of Σ. How large should p be such that the proportion of variance explained by the last (population) principal component account for less than 20% of total variation of X?

3.1 Find the eigenvalues Σ

$$\begin{aligned}
\Sigma &= diag(\lambda_1, \ldots, \lambda_n) - \pi_0\pi^T \\
&= \pi_0 \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} + \pi_0^2 \underbrace{\begin{pmatrix} -1 & -1 & \ldots & -1 \\ -1 & -1 & \ldots & -1 \\ \vdots & \vdots & \vdots & \vdots \\ -1 & -1 & \ldots & -1 \end{pmatrix}}_{J_p} \\
&= \pi_0 I_p + \pi_0^2 J_p \\
&= \pi_0 I_p - \pi_0^2 A
\end{aligned}$$

where $A = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1 & 1 & \ldots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \ldots & 1 \end{pmatrix}$

At first we want to find the eigenvalues of the matrix A.We observe that

$$Av = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1 & 1 & \ldots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \ldots & 1 \end{pmatrix} \underbrace{\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}}_{v} = pv$$

so by definiton of eigenvalue and eigenvectors we can say that the eigenvalue of our matrix A is $\lambda_1 = p$. The rank$(A) = 1$ so λ_1 is the only eigenvalue different from 0. **Spectral theorem (for real symmetric matrices)** Let $A \in \mathbb{R}^{n\times n}$ be real and symmetric. Then 1. The eigenvalues of A are real . 2. A is diagonalizable. 3. There is an orthonormal basis of $\mathbb{R}^n$ consiting on the eigenvectors of A. In short, A may be orthonormaly diagonalized: $A = VDV^T$ where $V \in \mathbb{R}^{n\times n}$ is an orthonormal matrix of eigenvectors of A and D is a real diagonal matrix of eigenvalues of A. So by the **Spectral Theorem** we can say that $\exists Q \in \mathbb{R}^{p\times p}$ orthonormal matrix such that $A = QDQ^T$ where $D = \begin{pmatrix} p & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{pmatrix}$.

Now we find the eigenvalue of $\Sigma = \pi_0(I_p - \pi_0 A)$. We recover A with QDQ^T:

$$\begin{aligned}
\Sigma &= \pi_0(I_p - \pi_0 QDQ^T) \\
&= \pi_0(QI_pQ^T - \pi_0 QDQ^T) \\
&= \pi_0 Q(I_p - \pi_0 D)Q^T
\end{aligned}$$

Defintion Similar Matrix A matrix A is said similar to $B \iff \exists P$ invertible such that $A = P^{-1}BP$.

From this definition and by how we wrote Σ we can say that the matrices Σ and $T := \pi_0(I_p - \pi_0 D)$ are similar so we can proceed by finding the eigenvalues of T (since two similar matrices share the same eigenvalues).

$$
\begin{aligned}
|T - \lambda I_p| &= \left| \begin{pmatrix} \pi_0(1 - p\pi_0) - \lambda & & & \\ & \pi_0 - \lambda & & \\ & & \ddots & \\ & & & \pi_0 - \lambda \end{pmatrix} \right| = 0 \\
&= [\pi_0(1 - p\pi_0) - \lambda] \cdot (\pi_0 - \lambda)^{p-1} = 0
\end{aligned}
$$

So the eigenvalues of Σ are $\tilde{\lambda}_1 = \pi_0(1 - p\pi_0)$ and $\tilde{\lambda}_2 = \cdots = \tilde{\lambda}_p = \pi_0$.

3.2 In order to find p such that the proportion of the variance explained by the last principal component is less than 20% we need to find the smallest eigenvalue.

$$\tilde{\lambda} = (\pi_0(1 - \pi_0 p), \underbrace{\pi_0, \dots, \pi_0}_{p-1})$$

$$
\begin{aligned}
&\pi_0 > \pi_0(1 - \pi_0 p) \\
&1 > (1 - \pi_0 p) \\
&\pi_0 p > 0 \quad \forall p > 0
\end{aligned}
$$

So we have that the smallest eigen value $\tilde{\lambda}_p = \pi_0(1 - \pi_0 p)$. Now we find the total variance as the sum of the eigenvalues of Σ:

$$
\begin{aligned}
Var(X) &= (1 - \pi_0 p)\pi_0 + \pi_0(p - 1) = \\
&= \pi_0 - \pi_0^2 p + \pi_0 p - \pi_0 \\
&= \pi_0 p(1 - \pi_0)
\end{aligned}
$$

And using the formula $\frac{\tilde{\lambda}_p}{Var(X)}$ we set the proportion less then 20%

$$
\begin{aligned}
&\frac{\pi_0(1 - \pi_0 p)}{\pi_0 p(1 - \pi_0)} < \frac{1}{5} \quad 0 < \pi_0 < \frac{1}{p} \\
&\frac{(1 - \pi_0 p)}{p(1 - \pi_0)} - \frac{1}{5} < 0 \\
&\frac{5\pi(1 - \pi_0 p) - \pi_0 p(1 - \pi_0)}{5\pi_0 p(1 - \pi)} < 0 \\
&\frac{5\pi_0 - 5\pi_0^2 p - \pi_0 p + \pi_0^2 p}{5\pi_0 p(1 - \pi_0)} < 0 \\
&\frac{-\pi_0^2 - \pi_0 p + 5\pi_0}{5\pi_0 p(1 - \pi_0)} < 0 \\
&\frac{\pi_0 p(-4\pi_0 - 1) + 5\pi_0}{5\pi_0 p(1 - \pi_0)} < 0 \\
&p(-4\pi_0 - 1) + 5 < 0 \\
&p > \frac{5}{4\pi_0 + 1}
\end{aligned}
$$

If we want to get rid on the dependence on π_0 and get a lower bound we ought to analyze the worst case scenario that is when the first (p-1) explain little variance hence π_0 is close to zero and since the sum has to be one the last would assume the greatest value possible. Now if we let $\pi_0 \to 0$ we get a lower bound not depending on π_0 which is the following:

$$p > \frac{5}{\underbrace{4\pi_0}_{0} + 1} = 5$$

Point 4

Perform a simulation study with $p = 3$, $\pi_0 = 1/4$ and $N = 1000$ Monte Carlo samples of $n = 100$ multinomially distributed Y_i. For $X = (X_1, X_2, X_3)$, make a scatterplot of the N values of X_2 vs X_1 and sketch the ellipse corresponding to the contour of the (theoretical limiting) bivariate density of (X_1, X_2) which contains 95% probability.

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
##
## Caricamento pacchetto: 'ellipse'
##
##
## Il seguente oggetto è mascherato da 'package:graphics':
##
##     pairs
```

We firstly initialize the variables such that they are coherent with the text, then we sample from a uniform discrete random variable (since each outcome is equally likely with probability $\pi_0 = \frac{1}{4}$). After we codify Y_i with one-hot random vectors.($Y_i \sim \text{Multinomial}\left(1, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$). Once we sample $n = 100$ Y_i, we sum the columns to get the random variable which is distributed as a $\sum_{i=1}^{100} Y_i \sim \text{Multinomial}\left(100, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$, we repeat this process $N = 1000$ and we apply the CLT on this 1000 sample to get the random vector $\sqrt{n}(\hat{\pi} - \pi) \sim \mathcal{N}_4(\underline{0}, \text{diag}(\pi) - \pi\pi^T)$.

```r
#We set seed to standardize result
set.seed(4567)
p = 3
pi_0 = 1/4
N = 1000
n= 100

norm_sample<- matrix(NA, nrow=N, ncol=4)
for (i in (1:N)){
  # Matrix nx4 for codified vector inizialed as full of NAs
  Y_hot_i = matrix(NA, nrow=n, ncol=4)
  # Samples from a uniform discrete (not yet codified)
  Y_i = sample(4,n,replace =T)
  for (j in (1:n)){
      # This transforms the Yi (not yet codified) into the codified vector
      Y_hot_i[j,] = as.numeric(1:4 %in% Y_i[j])
```

```
    }
  M = Y_hot_i %>%
  colSums()
  norm_sample[i,]<-(M/n - pi_0)*sqrt(n)
}
X <- norm_sample[,1:3]
```

We then pick the first 3 columns to get X, we plot the first two, and draw the countor line for the theoretical bivariate distribution parameterized as such $\mathcal{N}_2\left(\underline{0}, \begin{bmatrix} \frac{3}{16} & -\frac{1}{16} \\ -\frac{1}{16} & \frac{3}{16} \end{bmatrix}\right)$

```
Sigma = matrix( c(3/16 , -1/16, -1/16,3/16) ,nrow=2 , ncol=2)
sigma.e <- eigen(Sigma)
#eigen vectors
P <- sigma.e$vectors
lambda <- sigma.e$values
mu <- c(0,0)
e1 <- P[,1]
e2 <- P[,2]
b <- e1[2]/e1[1]
a <- b*mu[1]+ mu[2]
b2 <- e2[2]/e2[1]
a2 <- b2*mu[1]+ mu[2]
```

```
cc<-sqrt(qchisq(0.95, 2))

x<-cbind(cc*sqrt(lambda[1])*e1+mu,
         -cc*sqrt(lambda[1])*e1+mu,
         cc*sqrt(lambda[2])*e2+mu,
         -cc*sqrt(lambda[2])*e2+mu)
x
```

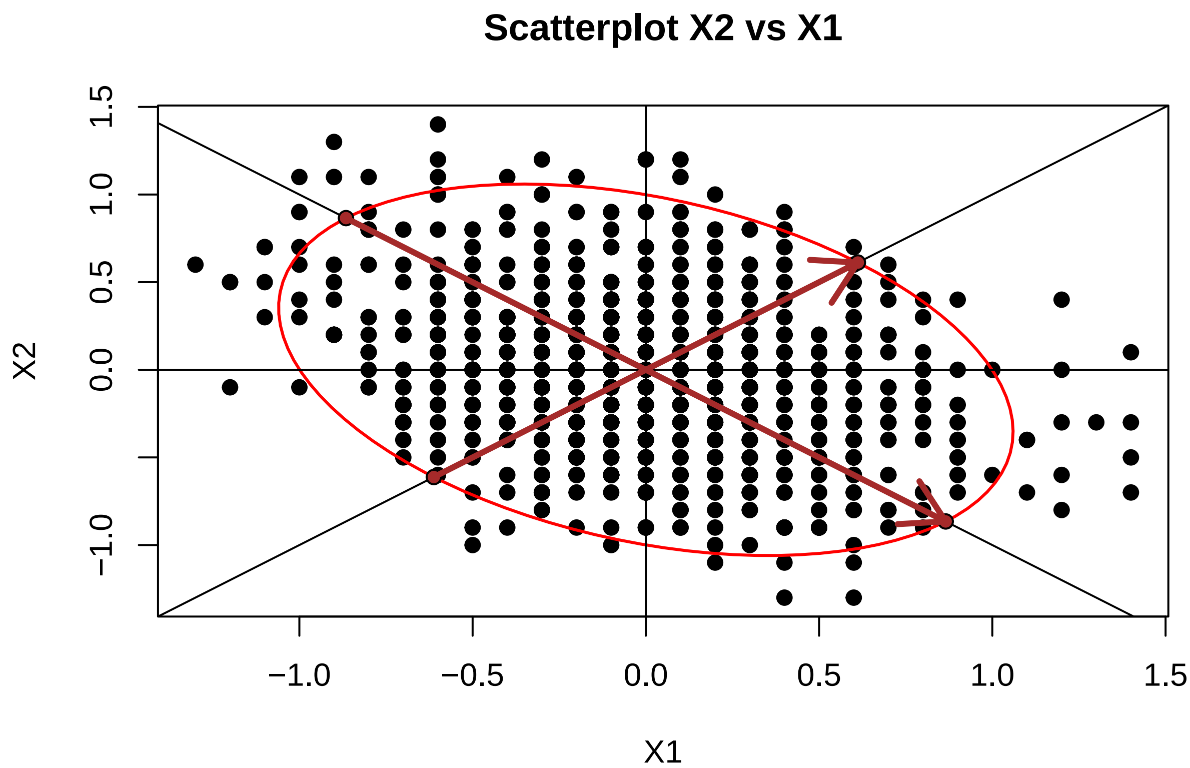
```
##             [,1]       [,2]       [,3]      [,4]
## [1,] -0.8654092  0.8654092 -0.6119367 0.6119367
## [2,]  0.8654092 -0.8654092 -0.6119367 0.6119367
```

```
mu = c(0,0)
Sigma = matrix( c(3/16 , -1/16, -1/16,3/16) ,nrow=2 , ncol=2)
Cor = matrix( c(1 , -1/3, -1/3,1) ,nrow=2 , ncol=2)
plot(X[,1],X[,2], xlab="X1", ylab = "X2", main="Scatterplot X2 vs X1",
     pch= 19)
lines(ellipse(x=Sigma,centre=mu,level=0.95),col="red",lwd=1.5)
abline(a = a , b=b)
abline(a= a2, b=b2)
abline(h=0,v=0)
points(x[1,],x[2,],pch=21,bg="brown")
x
```

```
##             [,1]       [,2]       [,3]      [,4]
## [1,] -0.8654092  0.8654092 -0.6119367 0.6119367
## [2,]  0.8654092 -0.8654092 -0.6119367 0.6119367
```

```
arrows(x[1,1],x[2,1],x[1,2],x[2,2],code=2,col="brown",lwd=3)

arrows(x[1,3],x[2,3],x[1,4],x[2,4],code=2,col="brown",lwd=3)
```




```r
solve(Sigma)
```

```
##      [,1] [,2]
## [1,]    6    2
## [2,]    2    6
```

```r
eigen(Sigma)
```

```
## eigen() decomposition
## $values
## [1] 0.250 0.125
##
## $vectors
##            [,1]       [,2]
## [1,] -0.7071068 -0.7071068
## [2,]  0.7071068 -0.7071068
```

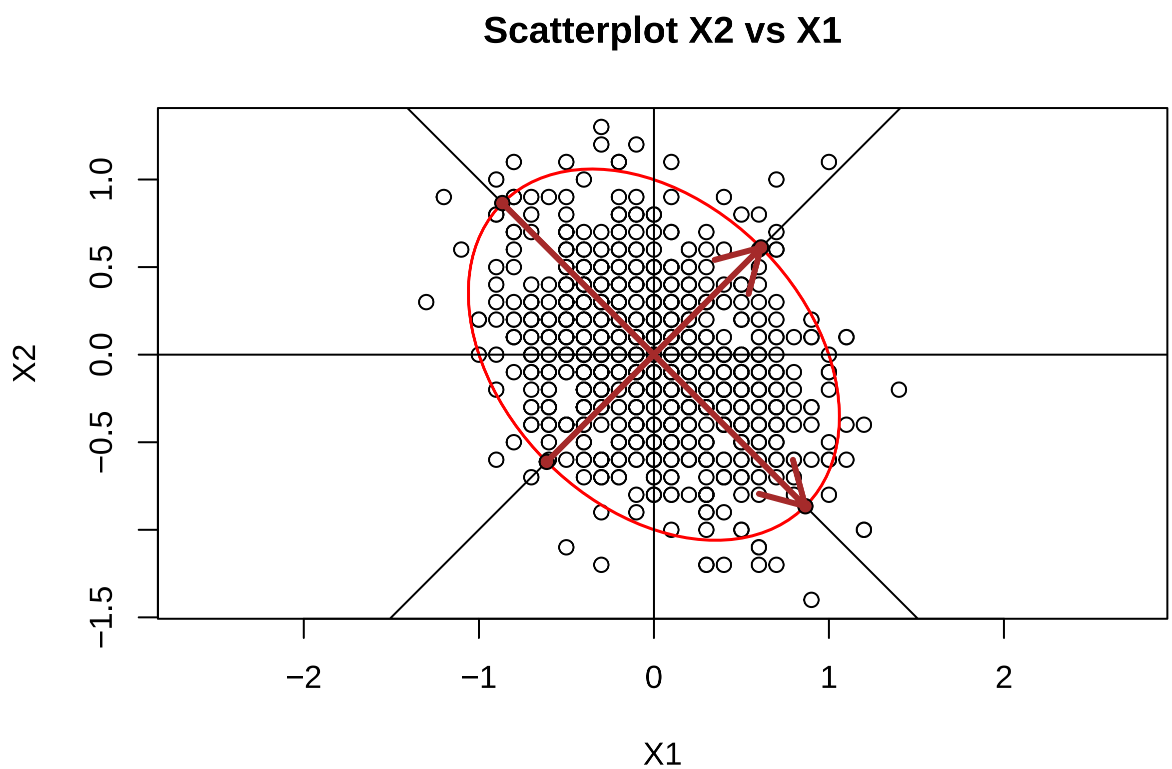
```r
set.seed(4567)

norm_samplev2<- matrix(NA, nrow=N, ncol=4)
for (i in (1:N)){
  M<-rmultinom(1,size=n,prob=c(1/4,1/4,1/4,1/4))
  norm_samplev2[i,]<-(M/n - pi_0)*sqrt(n)
}
Xv2 <- norm_samplev2[,1:3]
mu = c(0,0)
Sigma = matrix( c(3/16 , -1/16, -1/16,3/16) ,nrow=2 , ncol=2)
```

```r
plot(Xv2[,1],Xv2[,2], xlab="X1", ylab = "X2", main="Scatterplot X2 vs X1", asp=1)
lines(ellipse(x=Sigma,centre=mu,level=0.95),col="red",lwd=1.5)
abline(a = a , b=b)
abline(a= a2, b=b2)
abline(h=0,v=0)
points(x[1,],x[2,],pch=21,bg="brown")


arrows(x[1,1],x[2,1],x[1,2],x[2,2],code=2,col="brown",lwd=3)

arrows(x[1,3],x[2,3],x[1,4],x[2,4],code=2,col="brown",lwd=3)
```




```r
set.seed(4567)

M<-rmultinom(N,size=n,prob=c(1/4,1/4,1/4,1/4)) / n
norm_samplev2<-(M/n - pi_0)*sqrt(n)
cor(norm_sample)
```

```
##            [,1]       [,2]       [,3]       [,4]
## [1,]  1.0000000 -0.3292591 -0.3326393 -0.3230808
## [2,] -0.3292591  1.0000000 -0.3534907 -0.3402687
## [3,] -0.3326393 -0.3534907  1.0000000 -0.3208655
## [4,] -0.3230808 -0.3402687 -0.3208655  1.0000000
```

Point 5

Find the conditional distributions of $(X_1, X_2) \mid X_3 = x_3$ **and of** $X_3 \mid (X_1 = x_1, X_2 = x_2)$**.**

Given a Gaussian random vector we know that conditioning on it the conditioned vector is still Gaussian with parameters :

$$\begin{aligned} \overline{\mu} &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(a - \mu_2) \\ \overline{\Sigma} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \end{aligned}$$

We know that $(X_1, X_2, X_3) \sim \mathcal{N}_3(\underline{0}, \Sigma)$ where

$$\Sigma = \begin{pmatrix} \frac{3}{16} & -\frac{1}{16} & -\frac{1}{16} \\ -\frac{1}{16} & \frac{3}{16} & -\frac{1}{16} \\ -\frac{1}{16} & -\frac{1}{16} & \frac{3}{16} \end{pmatrix}$$

$$\left(\begin{array}{cc|c} \frac{3}{16} & -\frac{1}{16} & -\frac{1}{16} \\ -\frac{1}{16} & \frac{3}{16} & -\frac{1}{16} \\ \hline -\frac{1}{16} & -\frac{1}{16} & \frac{3}{16} \end{array}\right) = \left(\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right)$$

5.1 We need to find the distribution of $(X_1, X_2)|X_3 = x_3 \sim \mathcal{N}_2(\mu_{12|3}, \Sigma_{12|3})$.

So now we have to apply the formula above adapting the notation to our conditioned vector.

The mean vector is:

$$\begin{aligned} \mu_{12|3} &= \mu_{12} + \Sigma_{12}\Sigma_{22}^{-1}(x_3 - \mu_3) \\ &= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} -\frac{1}{16} \\ -\frac{1}{16} \end{pmatrix} \frac{16}{3}(x_3 - \mu_3) \\ &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -\frac{1}{3} \\ -\frac{1}{3} \end{pmatrix} x_3 = \begin{pmatrix} -\frac{1}{3}x_3 \\ -\frac{1}{3}x_3 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \Sigma_{12|3} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \\ &= \begin{pmatrix} \frac{3}{16} & -\frac{1}{16} \\ -\frac{1}{16} & \frac{3}{16} \end{pmatrix} - \begin{pmatrix} -\frac{1}{16} \\ -\frac{1}{16} \end{pmatrix} \frac{16}{3} \begin{pmatrix} -\frac{1}{16} & -\frac{1}{16} \end{pmatrix} \\ &= \begin{pmatrix} \frac{3}{16} & -\frac{1}{16} \\ -\frac{1}{16} & \frac{3}{16} \end{pmatrix} - \begin{pmatrix} \frac{1}{48} & \frac{1}{48} \\ \frac{1}{48} & \frac{1}{48} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{6} & -\frac{1}{12} \\ -\frac{1}{12} & \frac{1}{6} \end{pmatrix}. \end{aligned}$$

So a the end

$$(X_1, X_2)|X_3 = x_3 \sim \mathcal{N}_2\left(\begin{bmatrix} -\frac{1}{3}x_3 \\ -\frac{1}{3}x_3 \end{bmatrix}, \begin{bmatrix} \frac{1}{6} & -\frac{1}{12} \\ -\frac{1}{12} & \frac{1}{6} \end{bmatrix}\right).$$

5.2 We now have to find the gaussian random variable:

$$X_3|(X_1 = x_1, X_2 = x_2) \sim \mathcal{N}_1(\mu_{3|12}, \Sigma_{3|12}).$$

Now we get the mean:

$$
\begin{aligned}
\mu_{3|1,2} &= \mu_3 + \Sigma_{12}\Sigma_{11}^{-1}\begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix} \\
&= 0 + \begin{pmatrix} -\frac{1}{16} & -\frac{1}{16} \end{pmatrix} 32 \begin{pmatrix} \frac{3}{16} & \frac{1}{16} \\ \frac{1}{16} & \frac{3}{16} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\
&= \begin{pmatrix} -\frac{1}{16} & -\frac{1}{16} \end{pmatrix} \underbrace{\begin{pmatrix} 6 & 2 \\ 2 & 6 \end{pmatrix}}_{\Sigma_{11}^{-1}} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\
&= \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\
&= -\frac{1}{2}x_1 - \frac{1}{2}x_2
\end{aligned}
$$

And variance:

$$
\begin{aligned}
\Sigma_{3|12} &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \\
&= \frac{3}{16} - \begin{pmatrix} -\frac{1}{16} & -\frac{1}{16} \end{pmatrix} \begin{pmatrix} 6 & 2 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} -\frac{1}{16} \\ -\frac{1}{16} \end{pmatrix} \\
&= \frac{3}{16} - \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} -\frac{1}{16} \\ -\frac{1}{16} \end{pmatrix} \\
&= \frac{3}{16} - \frac{1}{16} = \frac{1}{8}
\end{aligned}
$$

So at the end:

$$
X_3|(X_1 = x_1, X_2 = x_2) \sim \mathcal{N}\left(-\frac{1}{2}x_1 - \frac{1}{2}x_2, \frac{1}{8}\right)
$$