GR5223 - HW6

Mathieu Sauterey - UNI: mjs2364

April 13, 2018

Problem 1

Over a 5-year period regular samples of fishermen on 28 lakes in Wisconsin were asked to report the time they spent fishing and how many of each type of game fish they caught. Their responses were then converted to a catch rate per hour for:

x_1 = Bluegill x_2 = Black crappie x_3 = Smallmouth bass x_4 = Largemouth bass

The estimated correlation matrix given below is based on a sample of about 120 (there were a few missing values).

```
# Creates the sample correlation matrix given by the problem
R <- matrix(c(1,0.4919,0.2636,0.4653, 0.4919, 1, 0.3127,0.3506,
              0.2636, 0.3127, 1, 0.4108, 0.4653, 0.3506, 0.4108, 1),
            nrow = 4, ncol = 4)
R
```

```
##        [,1]   [,2]   [,3]   [,4]
## [1,] 1.0000 0.4919 0.2636 0.4653
## [2,] 0.4919 1.0000 0.3127 0.3506
## [3,] 0.2636 0.3127 1.0000 0.4108
## [4,] 0.4653 0.3506 0.4108 1.0000
```

(a) Obtain the principal component solution for a factor model with k = 1.

```
# First, we perform a spectral decomposition on R to get its evalues and evectors
Gamma <- eigen(R)$vectors
Lambda <- diag(eigen(R)$values)

# Then we calculate Q_hat for k = 1 (one common factor)
Q_hat <- (Gamma %*% sqrt(Lambda) )[,1]
Q_hat
```

```
## [1] -0.7727495 -0.7386583 -0.6498683 -0.7673693
```

```
# We calculate Psi_hat (uniquenesses)
Psi_hat <- diag(diag(R - Q_hat %*% t(Q_hat)))
Psi_hat
```

```
##           [,1]     [,2]      [,3]      [,4]
## [1,] 0.4028583 0.000000 0.0000000 0.0000000
## [2,] 0.0000000 0.454384 0.0000000 0.0000000
## [3,] 0.0000000 0.000000 0.5776712 0.0000000
## [4,] 0.0000000 0.000000 0.0000000 0.4111444
```

```
R - (Q_hat %*% t(Q_hat) + Psi_hat)
```

```
##             [,1]        [,2]       [,3]       [,4]
## [1,]  0.00000000 -0.07889777 -0.2385854 -0.1276842
```

```
## [2,] -0.07889777  0.00000000 -0.1673306 -0.2162237
## [3,] -0.23858540 -0.16733061  0.0000000 -0.0878890
## [4,] -0.12768420 -0.21622367 -0.0878890  0.0000000
```

The one-factor model appears to fit the data well.

(b) Obtain the maximum likelihood solution for a factor model with k = 1. Are the principal component and maximum likelihood solutions consistent with each other?

```
# The function below performs factor extraction using the MLE method for k=1
factanal(covmat=R, factors=1, rotation="none")
```

```
##
## Call:
## factanal(factors = 1, covmat = R, rotation = "none")
##
## Uniquenesses:
## [1] 0.498 0.603 0.764 0.574
##
## Loadings:
##      Factor1
## [1,] 0.708
## [2,] 0.630
## [3,] 0.485
## [4,] 0.653
##
##                Factor1
## SS loadings      1.561
## Proportion Var   0.390
##
## The degrees of freedom for the model is 2 and the fit was 0.0571
```

The one-factor MLE model adequately explains 39% of the total sample variance. From this MLE extraction, we obtain the following loading and uniqueness estimates:

$\hat{\psi}_{MLE}$ = (0.498, 0.603, 0.764, 0.574)'

$\hat{Q}_{MLE}$ = (0.708, 0.630, 0.485, 0.653)'

We recall that the principal component (PC) solution was $\hat{Q}_{PC}$ = (-0.773, -0.739, -0.650, -0.767)'. The factor loadings from the principal component and maximum likelihood solutions are consistent with each other. Although their signs are inverted, the relative magnitude of the loadings within each solution is similar.

(c) Obtain the principal component solution for a factor model with k = 2, and rotate your solution. Interpret each factor.

```
# We calculate Q_hat_k2 for k = 2 (two common factors)
Q_hat_k2 <- (Gamma %*% sqrt(Lambda) )[,1:2]
Q_hat_k2
```

```
##            [,1]      [,2]
## [1,] -0.7727495  0.4056871
## [2,] -0.7386583  0.3656331
```

```
## [3,] -0.6498683 -0.6730125
## [4,] -0.7673693 -0.1905248
```

```
# We calculate Psi_hat_k2 (uniquenesses)
Psi_hat_k2 <- diag(diag(R - Q_hat_k2 %*% t(Q_hat_k2)))
Psi_hat_k2
```

```
##           [,1]      [,2]      [,3]      [,4]
## [1,] 0.2382762 0.0000000 0.0000000 0.0000000
## [2,] 0.0000000 0.3206964 0.0000000 0.0000000
## [3,] 0.0000000 0.0000000 0.1247254 0.0000000
## [4,] 0.0000000 0.0000000 0.0000000 0.3748447
```

```
# Uses Varimax to calculate an orthogonal rotation matrix of Q_hat2_k2
G <- varimax(Q_hat_k2)$rotmat
G
```

```
##            [,1]      [,2]
## [1,]  0.7824503 0.6227130
## [2,] -0.6227130 0.7824503
```

```
# Calculate the degree of the varimax rotation
acos(G[1,1])*180/pi
```

```
## [1] 38.51452
```

```
# Calculates the orthogonal rotation of Q_hat2_k2
Q_star <- Q_hat_k2 %*% G
Q_star
```

```
##             [,1]       [,2]
## [1,] -0.85726473 -0.1637711
## [2,] -0.80564787 -0.1738824
## [3,] -0.08939609 -0.9312803
## [4,] -0.48178608 -0.6269270
```

The factor loadings under the orthogonal rotation are printed above. The rotation is about 38.5 degrees counterclockwise.

The first factor can be interpreted as driving much of components 1, 2 (and somewhat of 4), respectively the hourly catch rates of Bluegill, Black crappie (and Largemouth bass).

The second factor can be interpreted as driving much of components 3 and 4, respectively the hourly catch rates of Smallmouth Bass and Largemouth bass.

Problem 2

The data file AirPollution.csv contains n = 42 measurements on air pollution variables recorded at 12:00 noon in the Los Angeles area on different days:

x_1 = Wind x_2 = Solar radiation $x_3 = NO_2$ $x_4 = O_3$

Compute the sample correlation matrix.

```
# Reads the data file and stores the dataset into a dataframe
data_air <- read.csv("AirPollution.csv")

# Computes the sample correlation matrix
```

```r
R2 <- cor(data_air)
R2
```

```
##            x1         x2         x3         x4
## x1  1.0000000 -0.1014419 -0.1098249 -0.2535928
## x2 -0.1014419  1.0000000  0.1157320  0.3191237
## x3 -0.1098249  0.1157320  1.0000000  0.1666422
## x4 -0.2535928  0.3191237  0.1666422  1.0000000
```

(a) Obtain the principal component solution to a factor model with k = 1.

```r
# First, we perform a spectral decomposition on R to get its evalues and evectors
Gamma2 <- eigen(R2)$vectors
Lambda2 <- diag(eigen(R2)$values)

# Then we calculate Q_hat2 for k = 1 (one common factor)
Q_hat2 <- (Gamma2 %*% sqrt(Lambda2) )[,1]
Q_hat2
```

```
## [1]  0.5638413 -0.6450049 -0.4769735 -0.7708354
```

```r
# We calculate Psi_hat2 (uniquenesses)
Psi_hat2 <- diag(diag(R2 - Q_hat2 %*% t(Q_hat2)))
Psi_hat2
```

```
##           [,1]      [,2]      [,3]      [,4]
## [1,] 0.6820829 0.0000000 0.0000000 0.0000000
## [2,] 0.0000000 0.5839686 0.0000000 0.0000000
## [3,] 0.0000000 0.0000000 0.7724963 0.0000000
## [4,] 0.0000000 0.0000000 0.0000000 0.4058129
```

```r
R2 - (Q_hat2 %*% t(Q_hat2) + Psi_hat2)
```

```
##           x1         x2         x3         x4
## x1 0.0000000  0.2622385  0.1591125  0.1810360
## x2 0.2622385  0.0000000 -0.1919183 -0.1780689
## x3 0.1591125 -0.1919183  0.0000000 -0.2010258
## x4 0.1810360 -0.1780689 -0.2010258  0.0000000
```

Thus, from the above 1-factor principal component extraction, we obtain the follwing estimates:

$\hat{Q}_{PC} = (0.564, -0.645, -0.477, -0.770)'$

$\hat{\psi}_{PC} = diag(0.682, 0.584, 0.772, 0.406)'$

(b) Find the maximum likelihood estimates of Q and Psi for k = 1.

```r
# The function below performs factor extraction using the MLE method for k=1
factanal(~x1+x2+x3+x4, data=data_air, factors=1, rotation="none")
```

```
##
## Call:
## factanal(x = ~x1 + x2 + x3 + x4, factors = 1, data = data_air,     rotation = "none")
##
## Uniquenesses:
```

```
##    x1    x2    x3    x4
## 0.895 0.832 0.946 0.405
##
## Loadings:
##    Factor1
## x1 -0.324
## x2  0.410
## x3  0.232
## x4  0.771
##
##                Factor1
## SS loadings      0.921
## Proportion Var   0.230
##
## Test of the hypothesis that 1 factor is sufficient.
## The chi square statistic is 0.15 on 2 degrees of freedom.
## The p-value is 0.93
```

The p-value suggests that the one-factor model adequately describes the data. It explains 23% of the total sample variance. From this above 1-factor MLE extraction, we obtain the following loading and uniqueness estimates:

$\hat{Q}_{MLE} = (-0.324, 0.410, 0.232, 0.771)'$

$\hat{\psi}_{MLE} = diag(0.895, 0.832, 0.946, 0.405)'$

(c) Compare the factorization obtained by the principal component and maximum likelihood methods.

From the 1-factor MLE extraction, we obtain the following loading estimates:

$\hat{Q}_{PC} = (0.564, -0.645, -0.477, -0.770)'$. This solution explains 39% of the total sample variance.

From the 1-factor principal component extraction, we obtain the following loading estimates:

$\hat{Q}_{MLE} = (-0.324, 0.410, 0.232, 0.771)'$. This solution explains 23% of the total sample variance.

The PC solution has a larger proportion of variance explained. The factor loadings from the principal component and maximum likelihood solutions seem to be consistent with each other. Although their signs are inverted, the relative magnitude of the loadings within each solution is fairly similar.

(d) Perform a varimax rotation of the principal component solution to a factor model with k = 2. Interpret the results.

```
# Then we calculate Q_hat2_k2 for k = 2 (two common factors)
Q_hat2_k2 <- (Gamma2 %*% sqrt(Lambda2) )[,1:2]
Q_hat2_k2
```

```
##            [,1]       [,2]
## [1,]  0.5638413 -0.2427710
## [2,] -0.6450049 -0.5212837
## [3,] -0.4769735  0.7351875
## [4,] -0.7708354 -0.1963048
```

```r
# We calculate Psi_hat2 (uniquenesses)
Psi_hat2_k2 <- diag(diag(R2 - Q_hat2_k2 %*% t(Q_hat2_k2)))
Psi_hat2_k2
```

```
##           [,1]      [,2]      [,3]      [,4]
## [1,] 0.6231452 0.0000000 0.0000000 0.0000000
## [2,] 0.0000000 0.3122319 0.0000000 0.0000000
## [3,] 0.0000000 0.0000000 0.2319956 0.0000000
## [4,] 0.0000000 0.0000000 0.0000000 0.3672773
```

```r
# Uses Varimax to calculate an orthogonal rotation matrix of Q_hat2_k2
G2 <- varimax(Q_hat2_k2)$rotmat
G2
```

```
##           [,1]       [,2]
## [1,] 0.8085335 -0.5884501
## [2,] 0.5884501  0.8085335
```

```r
# Calculate the degree of the varimax rotation
acos(G2[1,1])*180/pi
```

```
## [1] 36.0471
```

```r
# Calculates the orthogonal rotation of Q_hat2_k2
Q2_star <- Q_hat2_k2 %*% G2
Q2_star
```

```
##             [,1]        [,2]
## [1,]  0.31302600 -0.52808098
## [2,] -0.82825759 -0.04192211
## [3,]  0.04697214  0.87509884
## [4,] -0.73876180  0.29487917
```

The factor loadings under the orthogonal rotation are printed above. The rotation is about 36 degrees counterclockwise.

The first factor might reasonably be seen as some sort of contrast between Wind speed, and solar radiation and O_3 concentrations. The second factor might reasonably be seen as some sort of contrast between NO_2 and O_3 concentrations, and Wind speed.