

Statistisch Redeneren Week 4

Kwan Win Chung (10729585)

Matthijs Thoolen (10447822)

Younes Ouazref (10732519)

Mei 2016

Multivariate

19

Calculate the expectation and covariance of:

$$Z = \begin{pmatrix} X_1 + X_2 \\ X_1 - X_2 \end{pmatrix}$$

Given that: $\mu = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$

Expectation

$$Z = AX \text{ dus } \begin{pmatrix} X_1 + X_2 \\ X_1 - X_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

$$E(AX) = AE(X)$$

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} E\left(\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\right)$$

$$= \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} EX_1 \\ EX_2 \end{pmatrix}$$

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

De expectation is dan: $\begin{pmatrix} 0 \\ 2 \end{pmatrix}$

Covariance

Om de Covariance matrix te krijgen, maken we gebruik van: $\text{Cov}(AX) = A\text{Cov}(X)A^T$.

$$\text{Cov}(AX) = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$= \begin{pmatrix} 5 & -3 \\ -3 & 5 \end{pmatrix}$$

Dus kunnen we concluderen dat Z_1 en Z_2 independent zijn?

In de matrix zijn de niet diagonalen niet gelijk aan 0, met andere woorden ze zijn gecorreleerd. Hieruit kunnen we dus concluderen dat de componenten Z_1 en Z_2 independent zijn.

20

For which values of c are the components Z_1 en Z_2 uncorrelated?

Given: $X = (X_1 X_2)^T$ be a random vector with expectation $(00)^T$ (a zero mean random

covariance matrix) $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix}$ for $c > 0$.

Alles op een rijtje we hebben dus:

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, Z = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

$$\begin{aligned} \text{Cov}(BX) &= B\text{Cov}(X)B^T \\ &= \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 1+c & 1-c \\ 1-c & 1+c \end{pmatrix} \end{aligned}$$

We kunnen hieruit afleiden dat de componenten Z_1 en Z_2 uncorrelated zijn als c gelijk is aan 1.

21

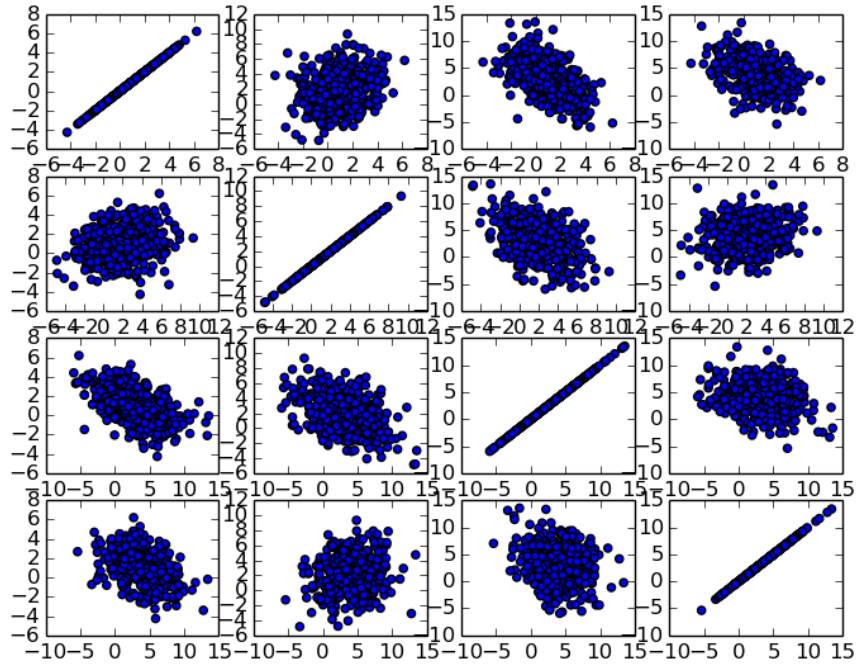


Figure 1: De 16 subplots

22

Voor opgave 22 hebben we dezelfde datamatrix van opgave 21 gebruikt als Sample.

$$\Sigma = \begin{bmatrix} 3.01602775 & 1.02746769 & -3.60224613 & -2.08792829 \\ 1.02746769 & 5.65146472 & -3.98616664 & 0.48723704 \\ -3.60224613 & -3.98616664 & 13.04508284 & -1.59255406 \\ -2.08792829 & 0.48723704 & -1.59255406 & 8.28742469 \end{bmatrix}$$

$$\mu = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

We gaan de functie nu met een steeds hogere N uitvoeren, naarmate N hoger is zal de nauwkeurigheid ook toenemen. De Python code is te vinden in het bestand "Opdracht_22.py".

```
('Mean (Estimated):', array([ 0.97021811,  1.73563458,  3.40085243,  3.74494214])) up in a
('Covariance matrix (Estimated):', array([[ 2.18908435,  0.43249669, -2.8555341 , -1.29283782],
      [ 0.43249669,  5.46187674, -3.72220448,  0.26699727],
      [-2.8555341 , -3.72220448, 13.60059608, -1.51178223],
      [-1.29283782,  0.26699727, -1.51178223,  7.07434389]]))
```

Figure 2: $N = 10^2$

```
('Mean (Estimated):', array([ 1.07566143,  2.01293263,  2.99037317,  3.9526706 ])) up in a
('Covariance matrix (Estimated):', array([[ 3.47975647,  1.18665815, -4.34782165, -2.46117975],
      [ 1.18665815,  5.93817165, -4.26592693,  0.135766 ],
      [-4.34782165, -4.26592693, 14.3679798 , -1.27005246],
      [-2.46117975,  0.135766 , -1.27005246,  8.89094402]]))
```

Figure 3: $N = 10^3$

```
('Mean (Estimated):', array([ 1.00572783,  2.0348016 ,  2.97552384,  4.03292677])) up in a
('Covariance matrix (Estimated):', array([[ 3.02889212,  1.0735162 , -3.64028456, -2.07453806],
      [ 1.0735162 ,  5.62775636, -4.05617835,  0.47112809],
      [-3.64028456, -4.05617835, 13.20153494, -1.68181689],
      [-2.07453806,  0.47112809, -1.68181689,  8.35141433]]))
```

Figure 4: $N = 10^4$

```
('Mean (Estimated):', array([ 0.99862033,  1.99957772,  2.99625915,  3.99995909])) up in a
('Covariance matrix (Estimated):', array([[ 3.01991335,  1.03390365, -3.61405447, -2.08447468],
      [ 1.03390365,  5.65545492, -3.99262599,  0.47649923],
      [-3.61405447, -3.99262599, 13.05082049, -1.56935508],
      [-2.08447468,  0.47649923, -1.56935508,  8.24826901]]))
```

Figure 5: $N = 10^6$

En inderdaad blijkt uit de experimenten dat naarmate N hoger is de nauwkeurigheid ook steeds verder toe neemt.

Nu moeten we Σ schatten, de data matrix bestaat alleen maar uit de geschatte μ . Volgens de theorie zouden we dan een Covariantie Matrix moeten vinden waarin alle elementen gelijk aan elkaar zijn. n is voor deze opgave fixed.

Nu voeren we de functie "calcMu()" uit, deze staat ook in "Opdracht_22.py".

```
('Mean (Estimated):', array([ 2.50392777,  2.50392777,  2.50392777,  2.50392777]))
('Covariance matrix (Estimated):', array([[ 1.36478855,  1.36478855,  1.36478855,  1.36478855],
      [ 1.36478855,  1.36478855,  1.36478855,  1.36478855],
      [ 1.36478855,  1.36478855,  1.36478855,  1.36478855],
      [ 1.36478855,  1.36478855,  1.36478855,  1.36478855]]))
```

Figure 6: data matrix bestaande uit enkel μ

Als we kijken naar de laatste afbeelding dan blijkt inderdaad de Covariantie Matrix enkel te bestaan uit dezelfde elementen. Dat klopt dus met de theorie.

PCA

5.1

Prove that the expression for the covariance matrix can be rewritten as:

$$\begin{aligned}
 S &= \frac{(\sum_{i=1}^n x_i x_i^T) - n m m^T}{n-1} \\
 \textbf{Given: } S &= \frac{1}{n-1} \sum_{i=1}^n (x_i - m)(x_i - m)^T \textbf{ and } m = \frac{1}{n} \sum_{i=1}^n x_i \\
 S &= \frac{1}{n-1} \sum_{i=1}^n (x_i - m)(x_i - m)^T \\
 &= \frac{1}{n-1} \sum_{i=1}^n (x_i x_i^T - x_i m^T - x_i^T m + m m^T) \\
 &= \frac{1}{n-1} (\sum_{i=1}^n (x_i x_i^T) + \sum_{i=1}^n (-x_i m^T - x_i^T m + m m^T)) \\
 &= \frac{1}{n-1} \sum_{i=1}^n ((x_i x_i^T) + (-m^T \sum_{i=1}^n x_i - m \sum_{i=1}^n x_i^T + n m m^T)) \\
 &= \frac{1}{n-1} \sum_{i=1}^n ((x_i x_i^T) + (-m^T n m - m n m^T + n m m^T)) \\
 &= \frac{1}{n-1} \sum_{i=1}^n ((x_i x_i^T) - n(m^T m + m m^T - m m^T)) \\
 &= \frac{(\sum_{i=1}^n x_i x_i^T) - n m m^T}{n-1}
 \end{aligned}$$

5.3

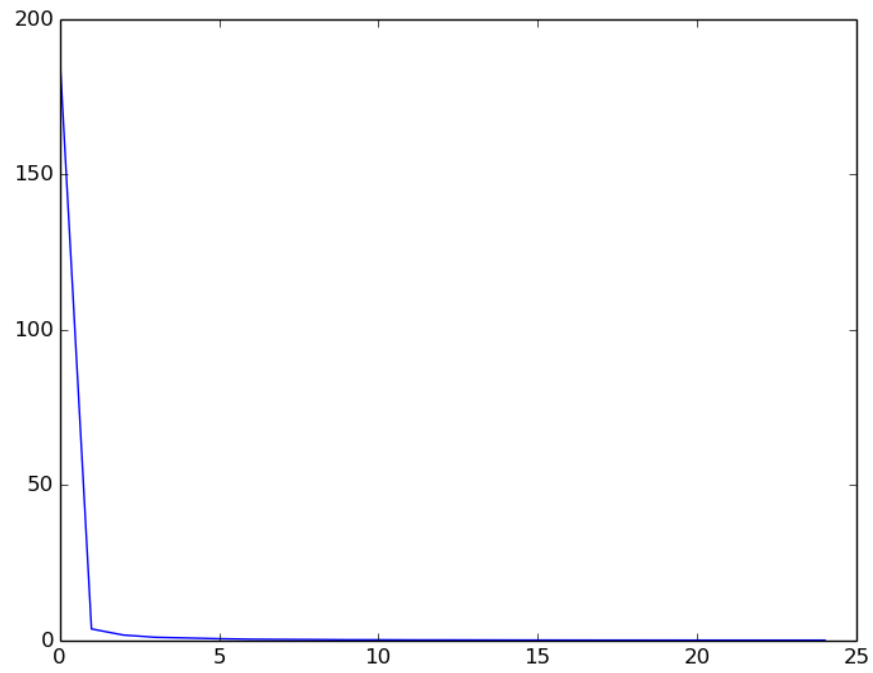


Figure 7: De scree diagram die de eigenvalues laat zien.