STT 843 Key to Homework 1 Spring 2018

Due date: Feb. 14, 2018

4.2. (a) Because $\sigma_{11} = 2$, $\sigma_{22} = 1$ and $\rho_{12} = 0.5$, we have $\sigma_{12} = \rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} = \sqrt{2}/2$. Then, the mean and covariance of the bivariate normal is

$$\mu = \begin{pmatrix} 0 \\ 2 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 2 & \sqrt{2}/2 \\ \sqrt{2}/2 & 1 \end{pmatrix}.$$

Then it follows that $|\Sigma| = 3/2$, $|\Sigma|^{1/2} = \sqrt{3/2}$ and

$$\Sigma^{-1} = \frac{2}{3}\begin{pmatrix} 1 & -\sqrt{2}/2 \\ -\sqrt{2}/2 & 2 \end{pmatrix}.$$

As a result, the bivariate normal density is

$$\begin{aligned} f(x) &= \frac{1}{(2\pi)\sqrt{3/2}} \exp\Big\{-\frac{1}{2}(x_1, x_2 - 2)\Sigma^{-1}\begin{pmatrix} x_1 \\ x_2 - 2 \end{pmatrix}\Big\} \\ &= \frac{1}{\sqrt{6}\pi} \exp\Big\{-\frac{1}{3}(x_1^2 - \sqrt{2}x_1(x_2 - 2) + 2(x_2 - 2)^2)\Big\}. \end{aligned}$$

(b) From part (a), we see that

$$(x - \mu)^T\Sigma^{-1}(x - \mu) = \frac{2}{3}(x_1^2 - \sqrt{2}x_1(x_2 - 2) + 2(x_2 - 2)^2).$$

(c) The constant density contours satisfies the equation: $(x-\mu)^T\Sigma^{-1}(x-\mu) = c$. The area that this contour contains is described by

$$R(c) = \{(x_1, x_2) : (x - \mu)^T\Sigma^{-1}(x - \mu) \leq c\}.$$

To find the constant density contours that contains 50% probability, we need to determine c such that $P\{X \in R(c)\} = 0.5$. This is equivalent to find c such that

$$P\{(X_1, X_2) : (X - \mu)^T\Sigma^{-1}(X - \mu) \leq c\} = 0.5.$$

Because $(X-\mu)^T\Sigma^{-1}(X-\mu)$ follows a chi-square distribution with degrees of freedom 2, the constant c should be the 50% quantile of the chi-square distribution. That is $c = \chi^2_{2,0.5}$. Thus, the contour that contains 50% probability is

$$(x - \mu)^T\Sigma^{-1}(x - \mu) = \chi^2_{2,0.5} = 1.386294.$$

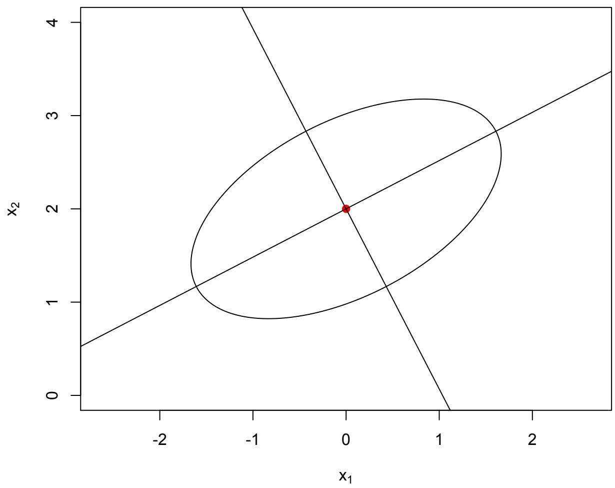

Figure 1: The density contour plot that contains 50% probability

4.4. (a) Let $c = (3, -2, 1)'$. Then, $3X_1 - 2X_2 + X_3 = c'X$ is linear combination of normally distributed random vector. Thus, $c'X$ has a normal distribution with mean $c'\mu$ and variance $c'\Sigma c$. Here $c'\mu = 3\times 2 + (-2)\times(-3) + 1\times 1 = 13$, and

$$c'\Sigma c = (3, -2, 1)\begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{pmatrix}\begin{pmatrix} 3 \\ -2 \\ 1 \end{pmatrix} = 9.$$

(b) Let $a = (a_1, a_2)'$. Because X_2 and $X_2 - a_1X_1 - a_2X_3$ are both normally distributed, X_2 and $X_2 - a_1X_1 - a_2X_3$ are independent if the covariance between X_2 and $X_2 - a_1X_1 - a_2X_3$ is 0. Then,

$$\mathrm{Cov}(X_2, X_2 - a_1X_1 - a_2X_3) = \sigma_{22} - a_1\sigma_{21} - a_2\sigma_{23} = 3 - a_1 - 2a_2 = 0.$$

Any vector $a = (a_1, a_2)'$ satisfying condition $3 - a_1 - 2a_2 = 0$ ensures the independence between X_2 and $X_2 - a_1X_1 - a_2X_3$.

4.6. (a) Because $\sigma_{12} = 0$, X_1 and X_2 are independent.

(b) Because $\sigma_{13} = -1$, X_1 and X_3 are not independent.

(c) Because $\sigma_{23} = 0$, X_2 and X_3 are independent.

(d) Because $(\sigma_{12}, \sigma_{32}) = (0, 0)$, (X_1, X_3) and X_2 are independent.

(e) Because $\mathrm{Cov}(X_1, X_1 + 3X_2 - 2X_3) = \sigma_{11} + 3\sigma_{12} - 2\sigma_{13} = 4 - 2(-1) = 6$, X_1 and $X_1 + 3X_2 - 2X_3$ are not independent.

4.7. (a) The joint distribution of $(X_1, X_3)'$ is normal with mean $(1,2)'$ and covariance $\begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix}$. By Result 4.6 in the textbook, we know that X_1 given $X_3 = x_3$ is normally distributed with mean $\mu_1 + \sigma_{13}\sigma_{33}^{-1}(x_3 - \mu_3) = 1 + (-1)2^{-1}(x_3 - 2) = -0.5x_3 + 2$ and covariance $\sigma_{11} - \sigma_{13}\sigma_{33}^{-1}\sigma_{31} = 4 - (-1)^2/2 = 3.5$.

(b) The conditional distribution of X_1 given $X_2 = x_2, X_3 = x_3$ is normal with mean

$$1 + (0, -1)\begin{pmatrix} 5 & 0 \\ 0 & 2 \end{pmatrix}^{-1}\begin{pmatrix} x_2 + 1 \\ x_3 - 2 \end{pmatrix} = -0.5x_3 + 1$$

and covariance

$$\sigma_{11} - (0, -1)\begin{pmatrix} 5 & 0 \\ 0 & 2 \end{pmatrix}^{-1}\begin{pmatrix} 0 \\ -1 \end{pmatrix} = 3.5.$$

4.16. (a) Because V_1 and V_2 are linear combinations of $X_1, \cdots, X_4$, V_1 and V_2 both have multivariate normal distribution. Therefore, we only need to determine the means and covariances of V_1 and V_2.

For V_1, the mean is

$$E(V_1) = \frac{1}{4}E(X_1) - \frac{1}{4}E(X_2) + \frac{1}{4}E(X_3) - \frac{1}{4}E(X_4) = 0$$

and covariance is

$$\mathrm{Cov}(V_1) = \frac{1}{16}\mathrm{Cov}(X_1) + \frac{1}{16}\mathrm{Cov}(X_2) + \frac{1}{16}\mathrm{Cov}(X_3) + \frac{1}{16}\mathrm{Cov}(X_4) = \frac{1}{4}\Sigma.$$

Therefore, V_1 is normally distributed with mean 0 and covariance $\frac{1}{4}\Sigma$.

For V_2, the mean is

$$E(V_2) = \frac{1}{4}E(X_1) + \frac{1}{4}E(X_2) - \frac{1}{4}E(X_3) - \frac{1}{4}E(X_4) = 0$$

and covariance is

$$\mathrm{Cov}(V_2) = \frac{1}{16}\mathrm{Cov}(X_1) + \frac{1}{16}\mathrm{Cov}(X_2) + \frac{1}{16}\mathrm{Cov}(X_3) + \frac{1}{16}\mathrm{Cov}(X_4) = \frac{1}{4}\Sigma.$$

Therefore, V_2 is normally distributed with mean 0 and covariance $\frac{1}{4}\Sigma$.

(b) Because $(V_1', V_2')'$ are linear combinations of $X_1, \cdots, X_4$, $(V_1', V_2')'$ is multivariate normally distributed. From part (a), the mean is 0. The covariance between V_1 and V_2 is

$$\mathrm{Cov}(V_1, V_2) = \frac{1}{16}\mathrm{Cov}(X_1) - \frac{1}{16}\mathrm{Cov}(X_2) - \frac{1}{16}\mathrm{Cov}(X_3) + \frac{1}{16}\mathrm{Cov}(X_4) = 0.$$

Therefore, the joint distribution of $(V_1', V_2')'$ is normal with mean 0 and covariance

$$\begin{pmatrix} \frac{1}{4}\Sigma & 0 \\ 0 & \frac{1}{4}\Sigma \end{pmatrix}.$$

As a result, the joint density of $(V_1', V_2')'$ is

$$f(V_1, V_2) = (2\pi)^{-p}|\frac{1}{4}\Sigma|^{-1} \exp\{-\frac{1}{2}V_1'(\frac{1}{4}\Sigma)^{-1}V_1 - \frac{1}{2}V_2'(\frac{1}{4}\Sigma)^{-1}V_2\}.$$

4.18. The maximum likelihood estimator for μ is

$$\bar{X}_n = \frac{1}{4}\sum_{i=1}^{4} X_i = \begin{pmatrix} 4 \\ 6 \end{pmatrix}.$$

The maximum likelihood estimator for Σ is

$$\hat{\Sigma} = \frac{1}{4}X'X - \bar{X}_n\bar{X}_n' = \begin{pmatrix} 1/2 & 1/4 \\ 1/4 & 3/2 \end{pmatrix}.$$

4.19. (a) The distribution of $(X_1 - \mu)'\Sigma^{-1}(X_1 - \mu)$ is chi-square distribution with degrees of freedom 6.

(b) The distribution of $\bar{X}$ is normal with mean μ and covariance $\Sigma/20$. The distribution of $\sqrt{n}(\bar{X} - \mu)$ is normal with mean 0 and covariance Σ.

(c) The distribution of $(n-1)S$ is Wishart distribution with degrees of freedom 19 and covariance Σ, namely $W_6(19, \Sigma)$.

4.21. (a) The distribution of $\bar{X}$ is normal with mean μ and covariance $\Sigma/60$.

(b) The distribution of $(X_1 - \mu)'\Sigma^{-1}(X_1 - \mu)$ is chi-square distribution with degrees of freedom 4.

(c) The distribution of $n(\bar{X} - \mu)'\Sigma^{-1}(\bar{X} - \mu)$ is chi-square distribution with degrees of freedom 4.

(c) The approximate distribution of $n(\bar{X} - \mu)'S^{-1}(\bar{X} - \mu)$ is chi-square distribution with degrees of freedom 4.

4.23. (a) The QQ plot is given in Figure 2. Based on the QQ plot, it seems that most data points are along the straight line. This might indicate that normal distribution is appropriate for this data set. However, due to the small sample size, the sample quantile estimated based on 10 data points might not be very informative. Therefore, the conclusion based on the QQ plot might not be reliable for this data set.

(b) Using the formula given in equation (4-31) on page 181, the correlation coefficient r_Q is defined by

$$r_Q = \frac{\sum_{j=1}^{n}(x_{(j)} - \bar{x})(q_{(j)} - \bar{q})}{\sqrt{\sum_{j=1}^{n}(x_{(j)} - \bar{x})^2}\sqrt{\sum_{j=1}^{n}(q_{(j)} - \bar{q})^2}}.$$

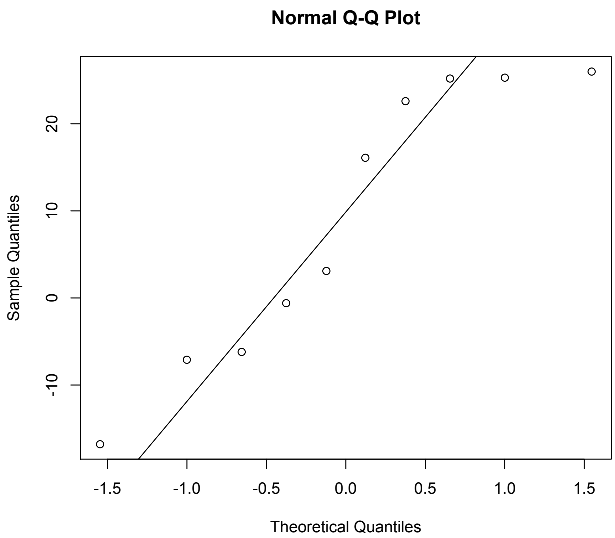

Figure 2: Q-Q plot of the annual rates of return given in question 4.23

The above coefficient r_Q is 0.9470451. The critical value for the test of normality can be found in Table 4.2. For our case, the sample size is 10 and significant level is 10%, the critical value is 0.9351. This means that r_Q is larger than the corresponding critical value. Thus, we fail to reject the null hypothesis and conclude that the normality assumption might be appropriate for this data set. We still need to be cautious about this conclusion because the data might not be independent. Therefore, the critical value used in our test might not be appropriate.

4.26. (a) Using R, we obtain the mean $\bar{x} = (5.200, 12.481)'$ and sample covariance as

$$\begin{pmatrix} 10.62222 & -17.71022 \\ -17.71022 & 30.85437 \end{pmatrix}.$$

Then the squared distances are

```
1.8753045 2.0203262 2.9009088 0.7352659 0.3105192
0.0176162 3.7329012 0.8165401 1.3753379 4.2152799
```

(b) Comparing the squared distances with the 50% quantile of chi-square distribution with degrees of freedom 2 (i.e., 1.386294), we find that the 4-th, 5-th, 6-th, 8-th and 9-th data points fall within the estimated 50% probability contour of a bivariate normal distribution. The proportion of the observations falling within the contour is 50%.

(c) A chi-square plot is given in the following Figure 4.

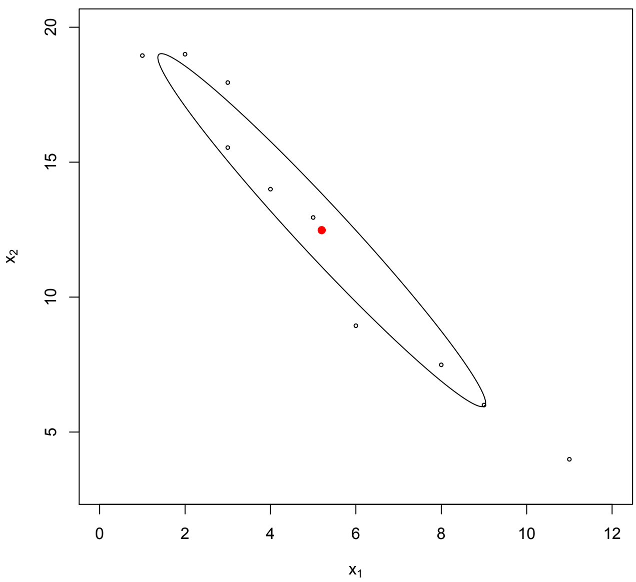

Figure 3: The density contour plot that contains 50% probability and scatter plot of data points

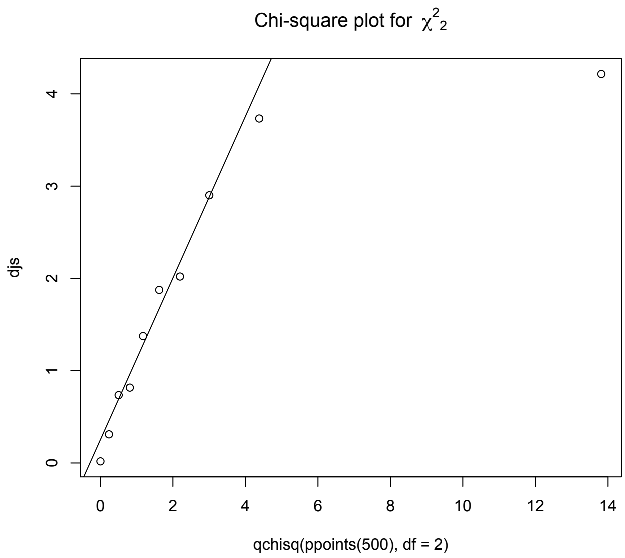

Figure 4: The chi-square plot for checking normality

(d) We observe from the chi-square plot given in part (c), most data points except the 10-th data point are along the theoretical line. Therefore, these data are approximately bivariate normal.