1

Consider a bivariate normal distribution with $\mu_1 = 1, \mu_2 = 3, \sigma_{11} = 2, \sigma_{22} = 1$ and $\rho_{12} = -0.8$

(a) Write out the bivariate normal density.

(b) write out the square statistics distance expression

$$(x-\mu)^T \Sigma^{-1} (x-\mu)$$

as a quadratic function of x_1 and x_2.

(c) Plot the constant-density contour that contains 50% of the probability

(a) 已知

$$\mu = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 2 & -0.8\sqrt{2} \\ -0.8\sqrt{2} & 1 \end{bmatrix}$$

透過計算可得

$$\Sigma^{-1} = \begin{bmatrix} 1.3889 & 1.5714 \\ 1.5714 & 2.7778 \end{bmatrix}$$

令 $x = (x_1, x_2)^T$，而透過定義可得 bivariate 的 density function 可寫成

$$f(x) = \frac{1}{2\pi|\Sigma^{\frac{1}{2}}|} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}, \quad -\infty < x_i < \infty, \quad i = 1, 2$$

(b) 承上題，透過計算

$$d^2 = \begin{bmatrix} x_1 - 1 & x_2 - 3 \end{bmatrix} \begin{bmatrix} 1.3889 & 1.5714 \\ 1.5714 & 2.7778 \end{bmatrix} \begin{bmatrix} x_1 - 1 \\ x_2 - 3 \end{bmatrix}$$

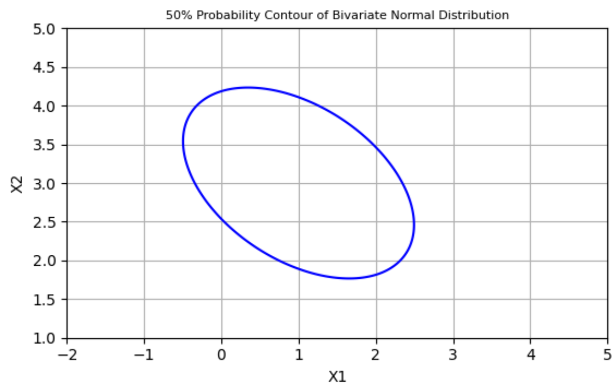
可得

$$d^2(x_1, x_2) = 2.9603(x_1-1)^2 + 3.1428(x_1-1)(x_2-3) + 4.3492(x_2-3)^2$$

(c) 已知 $(x-\mu)^T\Sigma^{-1}(x-\mu) \sim \chi_2^2$，因此可得

$$(x-\mu)^T\Sigma^{-1}(x-\mu) = \chi_2^2(0.5) = 1.386$$

透過 python 繪圖可得:



2

Let X be distributed as $N_3(\mu, \Sigma)$, where $\mu^T = (1, -1, 2)$ and

$$\Sigma = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix}$$

Which of the following random variables are independent? Explain

(a) X_1 and X_2
(b) X_1 and X_3
(c) X_2 and X_3
(d) (X_1, X_3) and X_2
(e) X_1 and $X_1 + 3X_2 - 2X_3$

(a) 我們可以透過 partition 得到 X_1, X_2 的 covariance matrix

$$\Sigma = \begin{bmatrix} 4 & 0 \\ 0 & 5 \end{bmatrix}$$

由於 X 服從 Multinormal distribution，且 $\sigma_{12} = 0$，可得 X_1, X_2 爲 independent

(b) 透過 partition 得到 X_1, X_3 的 covariance matrix

$$\Sigma = \begin{bmatrix} 4 & -1 \\ -1 & 2 \end{bmatrix}$$

$\sigma_{13} = -1 \neq 0$，因此 X_1, X_3 不爲 independent

(c) 透過 partition 得到 X_1, X_3 的 covariance matrix

$$\Sigma = \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix}$$

由於 X 服從 Multinormal distribution，且 $\sigma_{23} = 0$，可得 X_2, X_3 爲 independent

(d) 透過 partition 可得

$$\Sigma = \left[\begin{array}{c|c} \Sigma_{13} & \Sigma_{13,2} \\ \hline \Sigma_{2,13} & \Sigma_2 \end{array}\right] = \left[\begin{array}{cc|c} 4 & -1 & 0 \\ -1 & 2 & 0 \\ \hline 0 & 0 & 5 \end{array}\right]$$

由於 $\Sigma_{2,13} = \Sigma_{13,2} = 0$，因此 (X_1, X_3) 和 X_2 爲 independent

(e) 已知 $\sigma_{12} = 0, \sigma_{13} = -1$，因此

$$\begin{aligned} Cov(X_1, X_1 + 3X_2 - 2X_3) &= Cov(X_1, X_1) + 3Cov(X_1, X_2) - 2Cov(X_1, X_3) \\ &= \sigma_{11} + 3\sigma_{12} - 2\sigma_{13} \end{aligned}$$

$$= 4 + 3 \times 0 - 2 \times (-1)$$
$$= 6 \neq 0$$

因此 X_1 和 $X_1 + 3X_2 - 2X_3$ 不為 independent

3

Let $X_1, X_2, \cdots X_{20}$ be a random sample of size $n = 20$ from an $N_6(\mu, \Sigma)$ population. Specify each of the following completely.

(a) The distribution of $(X_1 - \mu)^T \Sigma^{-1} (X_1 - \mu)$

(b) The distribution of $\overline{X}$ and $\sqrt{n}(\overline{X} - \mu)$

(c) The distribution of $(n-1)S$

(a) 根據題意可得

$$(X_1 - \mu)^T \Sigma^{-1} (X_1 - \mu) \sim \chi^2_6$$

(b) 根據題意可得

$$\overline{X} \sim N_6(\mu, \frac{1}{20}\Sigma)$$
$$\sqrt{n}(\overline{X} - \mu) \sim N_6(0, \Sigma)$$

(c) 根據題意可得

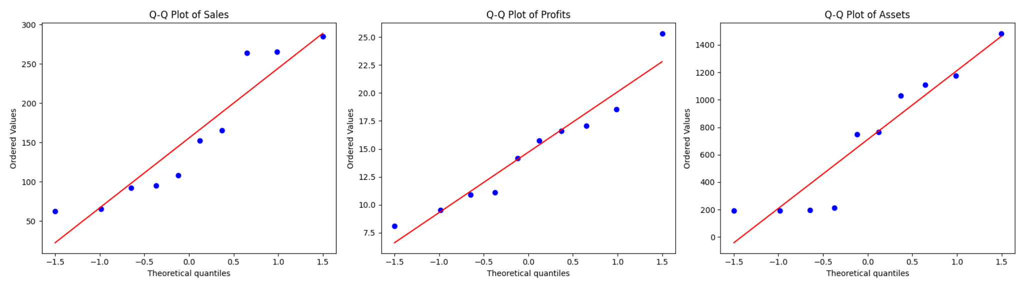
$$(n-1)S \sim W_6(19, \Sigma)$$ [1]

4

Exercise 1.4 contains data on three variables for the world's 10 largest companies as of April 2005. For the sales(x_1) and profit(x_2) data:

(a) Construct Q-Q plots. Do these data appear to be normally distributed? Explain

[1] Wishart distribution



透過觀察可發現三張圖的 sample 並沒有很好的 fit 圖中的紅線，因此可以合理推斷這三種變數並沒有服從常態分佈。

5

Exercise 1.2 gives the age x_1, measured in years, as well as the selling price x_2, measured in thousands of dollars, for $n = 10$ used cars. These data are reproduced as follows:

x_1	1	2	3	3	4	5	6	8	9	11
x_2	18.95	19.00	17.95	15.54	14.00	12.95	8.94	7.49	6.00	3.99

(a) Use the results of Exercise 1.2 to calculate the squared statistical distances $(\mathrm{x}_j - \bar{\mathrm{x}})^T \mathrm{S}^{-1} (\mathrm{x}_j - \bar{\mathrm{x}})$, where $\mathrm{x}_j^T = [x_{j1}, x_{j2}]$, $j = 1, 2, \ldots, 10$.

(b) Using the distances in Part a, determine the proportion of the observations falling within the estimated 50% probability contour of a bivariate normal distribution.

(c) Order the distances in Part a and construct a chi-square plot.

(a) 透過計算可得到 $\overline{x_1} = 5.2, \overline{x_2} = 12.481$，並且計算樣本共變異數矩陣

$$S = \begin{bmatrix} 10.62 & -17.71 \\ -17.71 & 30.85 \end{bmatrix}$$

根據定義

$$d^2 = (x_j - \overline{x})^T S^{-1} (x_j - \overline{x})$$

可求得

d_j^2	Statistical distance
d_1^2	1.36941757
d_2^2	1.4213818
d_3^2	1.70320545
d_4^2	0.85747649
d_5^2	0.55724247
d_6^2	0.13272605
d_7^2	1.93207174
d_8^2	0.90362611
d_9^2	1.17274803
d_{10}^2	2.05311468

(b) 已知

$$^{2}(x-\mu)^T\Sigma^{-1}(x-\mu) \sim \chi_p^2$$

且 $\chi_2^2(0.5) \approx 1.39$，因此可以整理出

d_j^2	Statistical distance	falling within the 50% contour
d_1^2	1.36941757	N
d_2^2	1.4213818	Y
d_3^2	1.70320545	Y
d_4^2	0.85747649	N
d_5^2	0.55724247	N
d_6^2	0.13272605	N
d_7^2	1.93207174	Y
d_8^2	0.90362611	N
d_9^2	1.17274803	N
d_{10}^2	2.05311468	Y

可知有 40% 的點落入 bivariate normal 的 50% contour，因此拒絕服從 bivariate normal 的假說

[2]Johnson, Richard claim that we can roughly expect the same percentage although we replace Σ^{-1} with S^{-1}

(c) 將 distance 由小至大排序並繪圖可得