

Practical 3: Bivariate Distributions in R

Main Aim: To recap **bivariate distributions** and the **bivariate Normal distribution** in R.

Today's practical is split into two halves:

1. **Simulating from bivariate distributions:** Directly related to the material in Chapter 6 of the course notes – two questions to try
2. **Mini project help:** We can give general support and guidance with the material in the mini project

Note that you should not attempt this practical unless you have completed all activities from week 6, including:

- Attend (or watch the recording of) the synchronous online lecture on Wednesday 2nd December
- Complete all work associated with Home Study video 8
- Complete all exercises in Chapter 6 of the course notes

Of course, you must also have access to RStudio. If you still don't have RStudio installed on your machine, visit our student resources webpage:

<http://www.ncl.ac.uk/maths-physics/students/resources/>

From this link, click on the link for our "Computing page", log-in using your University details, and then click on the link for "R". Here you will find some basic resources aimed at students who have never used R before; scroll down, and you'll find information about how to install R on your home PC. Alternatively, Chris Graham has posted instructions on installing RStudio to the stage 2 induction Canvas page.

Suggested study times have been attached to the activities below.

1 Bivariate Distributions

Instructions: Try out **all** of these questions relating to material in Chapter 6 of the course notes.

You will be taken to your breakout room to complete this work, in which you can ask your tutor (Lee/Amy/Kieran/Nicola) for any help you might need.

Suggested study time: Around 45 minutes.

1. Suppose that $X \sim N(100, 25)$ and that

$$f_{Y|X}(Y|X = x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-x)^2}.$$

Write an R function to sample realizations from (X, Y) , and explore the dependence between X and Y . [Hint: Do you recognize the distribution of $Y|X$? See Examples 6.1 and 6.2 in the lecture notes for more help]

2. Suppose

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 & -0.8\sqrt{2} \\ -0.8\sqrt{2} & 1 \end{pmatrix} \right).$$

- Write down the values of μ_x , μ_y , σ_x and σ_y .
- Find the correlation ρ .
- Find the conditional distribution for $Y|X = x$, and use this to simulate a sample of 1000 from the bivariate distribution for $(X, Y)^T$. Produce a plot of X versus Y , and use this to confirm your answer to part (b). [Hint: See result 2 on page 51 of the lecture notes, and Example 6.3]
- We will now repeat the simulation exercise in part (c) using the MASS package. Set up your mean vector and variance matrix with the following code:

```
1 mu=c(1,3)
2 sigma=matrix(data=c(2,-0.8*sqrt(2),-0.8*sqrt(2),1),nrow=2,byrow=TRUE)
```

Now install the MASS package, which can be used to simulate observations from a multivariate normal distribution; then explore the `mvrnorm` function:

```
1 library(MASS)
2 ?mvrnorm
```

Simulate a sample of 1000 pairs from the bivariate distribution for $(X, Y)^T$ using `mvrnorm`, and produce a plot of X versus Y . [Hint: See Example 6.4 in the lecture notes]

2 Mini Project

In the second half of today's practical we will be happy to provide general assistance with the mini project, due in by 4pm, this coming Thursday 17th December.

Solutions

1. Recognise that $Y|X = x \sim N(\mu = x, \sigma^2 = 1)$. Then:

```

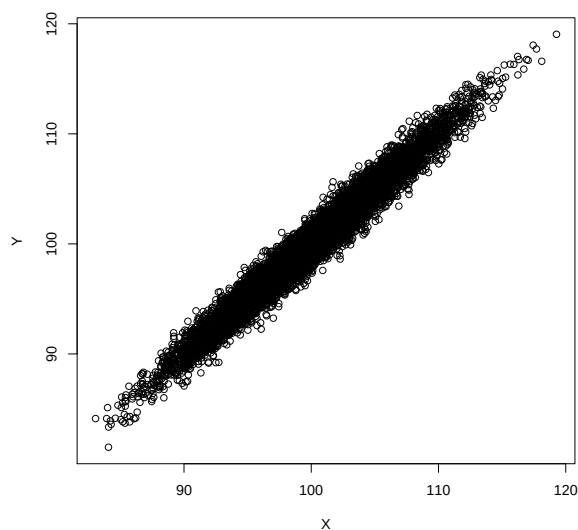
1 Q1 = function(N){
2   output = matrix(0, 2, N)
3   x = rnorm(N, 100, 5)
4   y = rnorm(N, x, 1)
5   output[1, ] = x
6   output[2, ] = y
7   return(output)}

```

```

1 test = Q1(1000)
2 plot(test[1,], test[2,], xlab="X", ylab="Y")
3 cor(test[1,], test[2,])
4 [1] 0.980935

```



2. (a) We have $\mu_x = 1$; $\mu_y = 3$; $\sigma_x = \sqrt{2}$; $\sigma_y = \sqrt{1}$
 (b) $\rho = \text{cov}(X, Y) / \sigma_x \sigma_y = -0.8\sqrt{2} / \sqrt{2}\sqrt{1} = -0.8$ (or just recognize from form of variance matrix)
 (c) From property (2) on page 25 of the lecture notes, we have

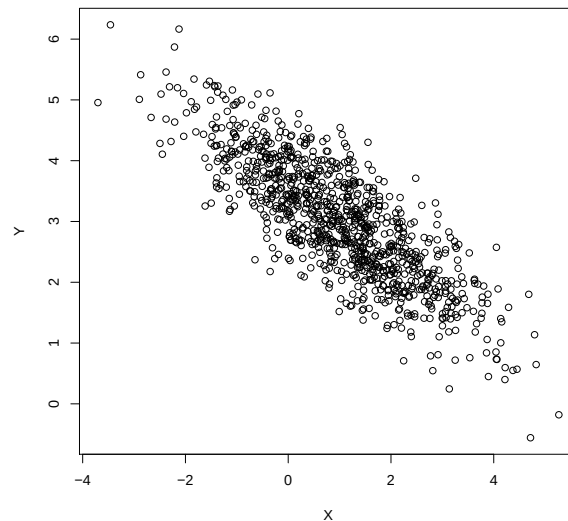
$$Y|X = x \sim N\left(3 + -0.8 \frac{\sqrt{1}}{\sqrt{2}}(x - 1), 1^2(1 - 0.64)\right) \equiv N(3 - 0.5656854(x - 1), 0.36).$$

Then:

```

1 Q2 = function(N){
2   output = matrix(0, 2, N)
3   x = rnorm(N, 1, sqrt(2))
4   y = rnorm(N, 3-0.5656854*(x-1), sqrt(0.36))
5   output[1, ] = x
6   output[2, ] = y
7   return(output)}
8
9 test = Q2(1000)
10 plot(test[1,], test[2,], xlab="X", ylab="Y")
11 cor(test[1,], test[2,])
12 [1] -0.8164408

```



```

1 library(MASS)
2 npts = 1000
3 mu = c(1, 3)
4 sigma = matrix(data = c(2, -0.8*sqrt(2), -0.8*sqrt(2), 1), nrow=2,
5                 byrow=TRUE)
6 test2 = mvrnorm(npts, mu, sigma)
7 plot(test2[,1], test2[,2])
8 cor(test2[,1], test2[,2])
9 [1] -0.7970329

```