Statistical Data Analysis
Problem sheet 2
Due Monday, 26 October 2020

Exercise 1: Suppose the joint pdf for the independent random variables x and y is given by

$$f(x,y) = \begin{cases} 1 & 0 < x < 1 \text{ and } 0 < y < 1 \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

(a) [6 marks] Show that the p.d.f. $g(z)$ for $z = xy$ is

$$g(z) = \begin{cases} -\ln z & 0 < z < 1 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

by defining an additional function u, which you can choose to be $u = x$. First, find the joint p.d.f. of z and u by using the Jacobian determinant as shown in the lecture notes. Integrate this over u to find the p.d.f. for z.

(b) [5 marks] Show that you obtain the same result for $g(z)$ using the formula for the Mellin convolution from the notes. In both (a) and (b) you will need to be careful about limits of integration.

Exercise 2 [4 marks]: Consider a random variable x and constants α and β. Starting from the definition of the expectation value for a continuous random variable, show that

$$\begin{aligned} E[\alpha x + \beta] &= \alpha E[x] + \beta\,, \\ V[\alpha x + \beta] &= \alpha^2 V[x]\,. \end{aligned} \tag{3}$$

Exercise 3: Consider two random variables x and y.

(a) [3 marks] Show that the variance of $\alpha x + y$ is given by

$$\begin{aligned} V[\alpha x + y] &= \alpha^2 V[x] + V[y] + 2\alpha \text{cov}[x,y] \\ &= \alpha^2 V[x] + V[y] + 2\alpha\rho\sigma_x\sigma_y\,, \end{aligned} \tag{4}$$

where α is any constant value, $\sigma_x^2 = V[x]$, $\sigma_y^2 = V[y]$, and the correlation coefficient is $\rho = \text{cov}[x,y]/\sigma_x\sigma_y$. (You may use the results from the previous exercise.)

(b) [3 marks] Using the result of (a), show that the correlation coefficient always lies in the range $-1 \leq \rho \leq 1$. (Use the fact that the variance $V[\alpha x + y]$ is always greater than or equal to zero and consider the cases $\alpha = \pm\sigma_y/\sigma_x$.)

Exercise 4 [5 marks]: Suppose the independent random variables x_1 and x_2 have means $\mu_1 = \mu_2 = 10$ and variances $\sigma_1^2 = \sigma_2^2 = 1$. Use error propagation to find the variance of

$$y = \frac{x_1^2}{x_2}\,. \tag{5}$$

Comment on the validity of the procedure if one had $\mu_2 = 1$.