DBS
Dublin Business School
excellence through learning

CA 1: probability modelling of data

Instructions:

- **All questions are mandatory**
- **Please upload R code with output in a zip file on Moodle**
- **The sharp deadline is 15 March 2018**

Q1: In a marketing system, there are three agents which report different types of data to the central manager. The first agent reports the number of customers entered to the sell zones per hour. The second agent records the time to arrival for service and the last agent reports whether the customer entered to the sell zone purchases the product or not.

(a) Specify the type of data reported by each agent and propose your probability model of data for each agent? (Hint: you might optionally select the parameters of your model) ***[15 marks]***

(b) If 10 customers enter to sell zones between 8:00-9:00 AM, and each customer purchases the product with p=0.8, what is the probability of this event that the reported value by the third agent is greater than 1? ***[10 marks]***

Please use R package

(c) Generate 100 samples from each agent using your proposed model in (a). [10 marks]

(d) Summarize and visualize the dataset of this system. ***[10 marks]***

Q2: Let the probability of the security of a wireless sensor is distributed as the beta model, Beta(2,4).

(a) What is your expectation of the security of this sensor? Find the variance? Let us assume that this sensor is a part of the network with 5

sensors. The network is stable if at least 2 sensors is secured. Based on this sensor, what is the probability of the security of the network? (Hint: use the expectation of beta model as the probability of the security of a sensor) ***[15 marks]***

Please use R package

(b) Generate 1000 samples from this network (200 samples from each sensor, i.e. $S_i \sim Beta(i, i+1), i = 1, \ldots, 5$). Find the sample mean and variance for all sensors. [10 marks]

(c) Which sensor is your preference in terms of the spread measure? Explain your reason.

(d) Sketch the pdf and cdf of the beta model in one page. ***[5 marks]***

Q3: Most graduate schools of business require applicants for admission to take the Graduate Management Admission Council's GMAT examination. Scores on the GMAT are roughly normally distributed with a mean of 500 and a standard deviation of 100. What is the probability of an individual scoring above 510 on the GMAT? (Verify your result using R) ***[10 marks]***

Please use R package

(a) Sketch 4 normal pdfs with different means and the same variance in one plot. ***[5 marks]***

(b) What is the probability of an individual scoring between 410 and 510 on the GMAT? ***[5 marks]***

Q4: In a big dataset, 4 features are extracted. Let us assume that the probability of the extraction of each feature is the same.

(a) What is the probability that we observe 8 data from the first feature where the sample size is n=10? ***[5 marks]***

(b) Compute the correlation between the first and third feature? ***[5 marks]***

Please use R package

Simulate the above dataset with 100 samples. Find the sample correlation between the first and the second features. ***[15 marks]***