COMP.4220 - Machine Learning

Homework 2

student name:

---

1. Suppose we roll a dice 10 times. Calculate the probability of rolling a 6 five times? What probability distribution did you use to explain this event? why? (1 Marks)

2. Suppose that the two variables x and z are statistically independent. Show that the mean and the variance of their sum satisfies

$$\mathbb{E}[x + z] = \mathbb{E}[x] + \mathbb{E}[z],$$
$$\mathbb{V}[x + z] = \mathbb{V}[x] + \mathbb{V}[z].$$

(3 Marks)

3. Show that the mean and the variance of the Binomial distribution can be written as

$$\mathbb{E}[m] = N\mu,$$
$$\mathbb{V}[m] = N\mu(1-\mu).$$

Hint: you can use the results of the previous question. (6 Marks)

4. Assume in our first trial, we gather N samples from a uniform distribution $\mathcal{U}[0, 1]$ and store the samples in vector s_1. In our second trial, we sample N more samples from the same distribution and this time store the average of the two sample sets as s_2. We continue this process until we reach s_{10}, in which we have stored the average of 10 sample sets from the same uniform distribution.
   (a) Write a `Python` script to replicate the above experiment for $N = 10,000$.
   (b) plot the histograms for s_1, s_2 and s_{10}. This must be a single plot including three subplots in a row with the range of x-axis limited to $[0, 1]$ and the range of y-axis limited to $[0, 5]$. Each subplot must have a legend or text that indicates s_1, s_2, or s_{10}. Also make sure the histogram represents a probability density function.
   (c) In a few sentences explain your observation about the results of averaging several sets of numbers sampled from a uniform distribution.

(15 Marks)