Assignment 2: CS 215

Due: 25th August before 11:55 pm

Remember the honor code while submitting this (and every other) assignment. All members of the group should work on all parts of the assignment. We will adopt a zero-tolerance policy against any violation.

Submission instructions:

1. You should type out all the answers to the written problems in Word (with the equation editor) or using Latex, or write it neatly on paper and scan it. In either case, prepare a pdf file.

2. Put the pdf file and the code for the programming parts all in one zip file. The pdf should contain the names and ID numbers of all students in the group within the header. The pdf file should also contain instructions for running your code. Name the zip file as follows: A2-IdNumberOfFirstStudent-IdNumberOfSecondStudent.zip. (If you are doing the assignment alone, the name of the zip file is A2-IdNumber.zip).

3. Upload the file on moodle BEFORE 11:55 pm on the due date (i.e. 25th August). We will nevertheless allow and not penalize any submission until 2:00 am on the following day (i.e. 26th August). No assignments will be accepted thereafter.

4. Note that only one student per group should upload their work on moodle.

5. Please preserve a copy of all your work until the end of the semester.

Questions:

1. Given random variables X and Y having probability density functions $f_X(x)$ and $f_Y(y)$ respectively and joint probability density function $f_{XY}(x,y)$, derive an expression for the probability density function of the random variable $Z = X+Y$ in terms of $f_X(x)$, $f_Y(y)$ and $f_{XY}(x,y)$. Also derive an expression for $P(X \leq Y)$. Refine the expressions if X and Y are independent. [5+5=10 points]

2. Read in the image parrots.png from the homework folder using the MATLAB function imread and cast it as a double array. Consider random variables X and Y where X denotes the pixel intensities from the image and Y denotes the intensities of the right neighbor of each pixel. Your task is to write MATLAB code to compute (1) the correlation coefficient, (2) a measure of dependence called quadratic mutual information (QMI) defined as $\sum_x \sum_y (p_{XY}(x,y) - p_X(x)p_Y(y))^2$, and (3) another measure of independence defined as $\sum_x \sum_y |p_{XY}(x,y) - p_X(x)p_Y(y)|$. Here $p_{XY}(x,y)$ represents the normalized joint histogram (*i.e.*, joint pmf) of X and Y ('normalized' means that the entries sum up to one). For X, ignore the first column of the image. For Y, ignore the last column of the image. For computing the joint histogram, use a bin-width of 10 in both X and Y. For computing the marginal histogram, you need to integrate the joint histogram along one of the two directions respectively. You should write your own joint histogram routine in MATLAB - do not use any inbuilt functions for it.

   Now randomly shuffle the pixel intensities of the image amongst the pixel locations using randperm and repeat the exercise. Now generate a uniform random noise image of the same size and having pixel intensities from 0 to 255. Repeat the exercise.

Your report should include a table containing all the three values for each case. Comment on your observations in your report. What is the minimum and maximum value that measure (3) can ever acquire? Can you place similar bounds for measure (2)? Explain in your report.

Marking scheme: 3 points for correlation coefficient routine, 4 points for joint histogram, 1 point for marginals, 1 point for measure (2), 1 point for measure (3), 3 points for comments on observations, 7 points for answering the questions about measure (2) and measure (3). [20 points]

3. Prove that independent random variables are uncorrelated (*i.e.*, their covariance is zero). Construct a counter-example to show that the converse is not always true. [5 + 5 = 10 points]

4. Consider X is a non-negative random variable. Prove that $P(E[X] \geq a)$ $P(X \geq a) \leq \dfrac{e^X}{e^a}$ where e is Napier's base. Also consider the function $Q_X(p)$ defined as the least value x for which $F_X(x) \geq p$. Then show that $Q_X(1 - p) \leq \dfrac{E(X)}{p}$. [5+5=10 points]

5. Consider two events A and B such that $P(A) = 0.6$ and $p(B) = 0.4$. What are the maximum and minimum values of $P(A \cap B)$ and $P(A \cup B)$? [5 points]

6. In class, we have seen the exclusive OR of events A and B denoted as $A \oplus B$, i.e. the event that exactly one of A or B occurs. Now consider a related problem. There is an island whose inhabitants speak one or more of n different languages. The proportion of people speaking those languages is $p_1, p_2, ..., p_n$ respectively where $\forall i, 0 \leq p_i \leq 1$. Determine the proportion of people who can speak exactly one language, assuming the ability to speak different languages are all independent. [5 points]