

# Assignment 2: CS 215

Due: 21st 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. 21st August). We will nevertheless allow and not penalize any submission until 2:00 am on the following day (i.e. 22nd 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  $Z = XY$  (i.e. the product of  $X$  and  $Y$ ) 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. Let  $X_1, X_2, \dots, X_n$  be  $n > 0$  independent identically distributed random variables with cdf  $F_X(x)$  and pdf  $f_X(x) = F'_X(x)$ . Derive an expression for the cdf and pdf of  $Y_1 = \max(X_1, X_2, \dots, X_n)$  and  $Y_2 = \min(X_1, X_2, \dots, X_n)$  in terms of  $F_X(x)$ . [10 points]
3. Consider a sequence of Bernoulli trials with success parameter  $p \in (0, 1)$  where you bet a certain amount. If a trial results in success, you gain that amount, else you pay the same amount. Now consider a strategy where you bet amount  $x$  in the first trial. If the trial is a failure, you bet double the amount in the next trial, and stop as soon as a trial is a success. Compute the total amount you will win if  $n$  trials are played (i.e. you won on the  $n^{\text{th}}$  trial). [10 points]
4. Prove that any two independent random variables are uncorrelated (i.e., their covariance is zero). Construct a counter-example to show that the converse is not always true. [2 + 5 = 7 points]
5. Given a random variable  $X$  whose values lie in the interval  $[a, b]$  where  $a < b$ , prove that  $\text{Var}(X) \leq \frac{(b-a)^2}{4}$ . [8 points]

6. A function  $g(x)$  of variable  $x$  is said to be convex if it always lies above its tangent. In other words if the line  $l(x) = a + bx$  is a tangent at  $x$ , then  $g(x) \geq a + bx$ . For example,  $g(x) = x^2$  is a convex function. Prove that  $g(E(X)) \geq E(g(X))$ . [10 points]
7. (a) 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  $X_1, X_2$  where  $X_1$  is a random variable that denotes the pixel intensities from the image and  $X_2$  denotes the intensities of a neighbor of each pixel located  $z$  pixels away. For the case  $z = 1$ , 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  $\sum_{x_1} \sum_{x_2} (p_{X_1 X_2}(x_1, x_2) - p_{X_1}(x_1)p_{X_2}(x_2))^2$ , and (3) another measure of independence defined as  $\sum_{x_1} \sum_{x_2} |p_{X_1 X_2}(x_1, x_2) - p_{X_1}(x_1)p_{X_2}(x_2)|$ . Here  $p_{X_1 X_2}(x_1, x_2)$  represents the normalized joint histogram (*i.e.*, joint pmf) of  $X_1$  and  $X_2$  ('normalized' means that the entries sum up to one). For  $X_2$ , ignore the first  $z$  columns of the image. For  $X_1$ , ignore the last  $z$  columns of the image. For computing the joint histogram, use a bin-width of 10 in both  $X_1$  and  $X_2$ . 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.
- (b) Compute only the correlation coefficient for different values of  $z$  ranging from 0 to 100. Plot a graph of the correlation coefficient versus  $z$ . Comment on the graph.
- Your report should include a table containing all the three values for each case. Comment on the differences in the values for the original and scrambled image in your report. What is the minimum and maximum value that measure (3) can ever acquire? Explain in your report. [15 points]