Assignment 2: CS 215

Due: 2nd September before 11:55 pm

All members of the group should work on all parts of the assignment. Copying across groups or from other sources is not allowed. We will adopt a zero-tolerance policy against any violation.

Submission instructions:

1. You should type out a report containing 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 single pdf file.
2. The report should contain names and roll numbers of all group members on the first page as a header.
3. 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-IdNumberofThirdStudent.zip. (If you are doing the assignment alone, the name of the zip file is A2-IdNumber.zip, if there are two students it should be A2-IdNumberOfFirstStudent-IdNumberOfSecondStudent.zip).
4. Upload the file on moodle BEFORE 11:55 pm on the due date. We will nevertheless allow and not penalize any submission until 10:00 am on the following day (i.e. 3rd September). No assignments will be accepted thereafter.
5. Note that only one student per group should upload their work on moodle, though all group members will receive grades.
6. Please preserve a copy of all your work until the end of the semester.

Questions:

1. Let $X_1, X_2, ..., X_n$ 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 $Y_1 = \max(X_1, X_2, ..., X_n)$ and $Y_2 = \min(X_1, X_2, ..., X_n)$ in terms of $F_X(x)$. [10 points]
Solution: $F_{Y_1}(x) = P(Y_1 \leq x) = P(X_1 \leq x, X_2 \leq x, ..., X_n \leq x) = \prod_{i=1}^{n} P(X_i \leq x) = \prod_{i=1}^{n} F_{X_i}(x) = (F_X(x))^n$. Hence $f_{Y_1}(x) = n(F_X(x))^{n-1} f_X(x)$.
$P(Y_2 \geq x) = P(X_1 \geq x, X_2 \geq x, ..., X_n \geq x) = \prod_{i=1}^{n} P(X_i \geq x) = (1 - F_X(x))^n$. Hence $F_{Y_2}(x) = 1 - P(Y_2 \geq x) = 1 - (1 - F_X(x))^n$. Hence $f_{Y_2}(x) = n(1 - F_X(x))^{n-1} f_X(x)$.

2. We say that a random variable X belongs to a Gaussian mixture model (GMM) if $X \sim \sum_{i=1}^{K} p_i \mathcal{N}(\mu_i, \sigma_i^2)$ where p_i is the 'mixing probability' for each of the K constituent Gaussians, with $\sum_{i=1}^{K} p_i = 1; \forall i, 0 \leq p_i \leq 1$. To draw a sample from a GMM, we do the following: (1) One of the K Gaussians is randomly chosen as per the PMF $\{p_1, p_2, ..., p_K\}$ (thus, a Gaussian with a higher mixing probability has a higher chance of being picked). (2) Let the index of the chosen Gaussian be (say) m. Then, you draw the value from $\mathcal{N}(\mu_m, \sigma_m^2)$. If X belongs to a GMM as defined here, obtain expressions for $E(X), \text{Var}(X)$ and the MGF of X.
Now consider a random variable of the form $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ for each $i \in \{1, 2, ..., K\}$. Define another random variable $Z = \sum_{i=1}^{K} p_i X_i$. Derive an expression for $E(Z), \text{Var}(Z)$ and the PDF, MGF of Z. [2+2+2+2+2+2+3=15 points]
Solution: We have $E(X) = \sum_{i=1}^{K} p_i E(X_i) = \sum_{i=1}^{K} p_i \mu_i$. We also have $E(X^2) = \sum_{i=1}^{K} p_i E(X_i^2) =$

$\sum_{i=1}^K p_i(\sigma_i^2+\mu_i^2)$. Hence $Var(X) = E(X^2) - [E(X)]^2 = \sum_{i=1}^K p_i(\sigma_i^2+\mu_i^2) - [\sum_{i=1}^K p_i\mu_i]^2$. The MGF is given by $\Phi_X(t) = E(e^{tX}) = \sum_{i=1}^K p_i E(e^{tX_i}) = \sum_{i=1}^K p_i e^{\mu_i t + \sigma_i^2 t^2/2}$.
The random variables in $\{X_i\}_{i=1}^K$ are independent. We have $\Phi_{p_iZ_i}(t) = \exp(p_i\mu_i t + p_i^2\sigma_i^2 t^2/2)$. Hence we have $\Phi_{p_iZ_i}(t) = \exp(\sum_{i=1}^K p_i\mu_i t + p_i^2\sigma_i^2t^2/2)$. This corresponds to a Gaussian $\mathcal{N}(\bar{\mu},\bar{\sigma}^2)$ where $\bar{\mu} = \sum_{i=1}^K p_i\mu_i, \bar{\sigma}^2 = \sum_{i=1}^K p_i^2\sigma_i^2$. Hence the PDF of Z is $f_Z(x) = \exp(-(x-\bar{\mu})^2/(2\bar{\sigma}^2))/(\bar{\sigma}\sqrt{2\pi})$. Also $E(Z) = \bar{\mu}, Var(Z) = \bar{\sigma}^2$.

3. Using Markov's inequality, prove the following one-sided version of Chebyshev's inequality for random variable X with mean μ and variance σ^2: $P(X-\mu \geq \tau) \leq \dfrac{\sigma^2}{\sigma^2+\tau^2}$ if $\tau > 0$, and $P(X-\mu\geq\tau) \geq 1 - \dfrac{\sigma^2}{\sigma^2+\tau^2}$ if $\tau < 0$. [15 points]
Solution: Consider the $\tau > 0$ case. Let $Y = X - \mu$. Now, $P(X-\mu\geq\tau) = P(Y\geq\tau) = P(Y+b\geq\tau+b)$. Now $P(Y+b\geq\tau+b) \leq P((Y+b)^2 \geq (\tau+b)^2) \leq \dfrac{E[(Y+b)^2]}{(\tau+b)^2} = \dfrac{\sigma^2+b^2}{(\tau+b)^2}$. We want to consider b which minimizes the RHS. One can show that this corresponding value of b is equal to σ^2/τ. Substitution yields $P(X-\mu\geq\tau) \leq \dfrac{\sigma^2}{\sigma^2+\tau^2}$ if $\tau > 0$.
For the other side when $\tau < 0$, consider $\gamma = -\tau > 0$. Then $P(X-\mu<\tau) = P(-Y>\gamma) \leq \dfrac{\sigma^2}{\gamma^2+\sigma^2}$ using the previous set of steps with $-Y$ and γ. Taking the complement set, we get $P(X-\mu\geq\tau) \geq 1 - \dfrac{\sigma^2}{\sigma^2+\tau^2}$.
Marking scheme: 7.5 points for each inequality. 5 points for expressing everything in terms of b and 2.5 points for the optimization. If the first derivation is right, steps from it can be used for the second one directly. If incorrect steps from the first derivation are used in the second, the student loses equal points in both places.

4. Given stuff you've learned in class, prove the following bounds: $P(X\geq x) \leq e^{-tx}\phi_X(t)$ for $t > 0$, and $P(X\leq x) \leq e^{-tx}\phi_X(t)$ for $t < 0$. Here $\phi_X(t)$ represents the MGF of random variable X for parameter t.
Now consider that X denotes the sum of n independent Bernoulli random variables $X_1, X_2, ..., X_n$ where $E(X_i) = p_i$. Let $\mu = \sum_{i=1}^n p_i$. Then show that $P(X > (1+\delta)\mu) \leq \dfrac{e^{\mu(e^t-1)}}{e^{(1+\delta)t\mu}}$ for any $t\geq 0, \delta > 0$. You may use the inequality $1 + x \leq e^x$. Further show how to tighten this bound by choosing an optimal value of t. [15 points]
Solution: Using Markov's inequality, we have $P(X\geq x) = P(e^{tX}\geq e^{tx}) \leq \dfrac{E(e^{tX})}{e^{tx}} = \phi_X(t)e^{-tx}$ when $t > 0$. When $t < 0$, we have $P(X\leq x) = P(e^{tX}\geq e^{tx}) \leq \dfrac{E(e^{tX})}{e^{tx}} = \phi_X(t)e^{-tx}$.
We have $P(X > (1+\delta)\mu) = P(e^{tX} > e^{(1+\delta)\mu} \leq \dfrac{E[e^{tX}]}{e^{(1+\delta)t\mu}}$. The numerator is equal to $E^{e^{t\sum_i X_i}} = E(\Pi_i e^{tX_i}) = \Pi_i E[e^{tX_i}] = \Pi_i(p_ie^t + 1 - p_i) = \Pi_i[1 + p_i(e^t-1)]$. Now, we substitute $1 + x \leq e^x$ and this gives us:
$E^{e^{tX}} \leq \Pi_i e^{p_i(e^t-1)} = e^{e^{\sum_i p_i(e^t-1)}} = e^{(e^t-1)\mu}$.
This finally gives us: $P(X > (1+\delta)\mu) \leq \dfrac{e^{\mu(e^t-1)}}{e^{(1+\delta)t\mu}}$.
For the last part, one can show that the value of t which minimizes the RHS is given by $t = \log(1+\delta)$. Substitution yields the tighter bound
$P(X > (1+\delta)\mu) \leq e^{(\delta-(1+\delta)\log(1+\delta))\mu}$.
Marking scheme: For the first two inequalities, 4 points - they are a direct usage of Markov's inequality. For the third inequality, which is the major one, there are 8 points: 2 points for initial Markov's inequality, 4 points for the RHS $\Pi_i[1+p_i(e^t-1)]$ and 2 points for the final step. For the refinement with optimal t, there are further 3 points.

5. Consider N independent random variables $X_1, X_2, ..., X_N$, such that each variable X_i takes on the values $1, 2, 3, 4, 5$ with probability $0.05, 0.4, 0.15, 0.3, 0.1$ respectively. For different values of $N \in \{5, 10, 20, 50, 100, 200, 500, 1000, 5000, 10000\}$, do as follows:

   (a) Plot the (empirically determined) distribution of the average of these random variables ($X_{avg}^{(N)} = \sum_{i=1}^N X_i/N$) in the form of a histogram with 50 bins.

(b) Empirically determine the CDF of $X_{avg}^{(N)}$ using the ecdf command of MATLAB (this is called the empirical CDF). On a separate figure, plot the empirical CDF. On this, overlay the CDF of a Gaussian having the same mean and variance as $X_{avg}^{(N)}$. To get the CDF of the Gaussian, use the normcdf function of MATLAB.

Solution: Note that the mean and variance of $X_{avg}^{(N)}$ are μ and σ^2/N respective where $E(X_i) = \mu, Var(X_i) = \sigma^2$ for any i. In this specific, case $\mu = 3, \sigma^2 = 1.3$ using the PMF.

(c) Let $E^{(N)}$ denote the empirical CDF and $\Phi^{(N)}$ denote the Gaussian CDF. Compute the maximum absolute difference (MAD) between $E^{(N)}(x)$ and $\Phi^{(N)}(x)$ numerically, at all values x returned by ecdf. For this, read the documentation of ecdf carefully. Plot a graph of MAD as a function of N. [3+3+4 = 10 points]

6. Read in the images T1.jpg and T2.jpg from the homework folder using the MATLAB function imread and cast them as a double array. These are magnetic resonance images of a portion of the human brain, acquired with different settings of the MRI machine. They both represent the same anatomical structures and are perfectly aligned (i.e. any pixel at location (x, y) in both images represents the exact same physical entity). Consider random variables I_1, I_2 which denote the pixel intensities from the two images respectively. Write a piece of MATLAB code to shift the second image along the X direction by t_x pixels where t_x is an integer ranging from -10 to +10. While doing so, assign a value of 0 to unoccupied pixels. For each shift, compute the following measures of dependence between the first image and the *shifted version* of the second image:

   - the correlation coefficient ρ,
   - a measure of dependence called quadratic mutual information (QMI) defined as $\sum_{i_1} \sum_{i_2} (p_{I_1 I_2}(i_1, i_2) - p_{I_1}(i_1) p_{I_2}(i_2))^2$, where $p_{I_1 I_2}(i_1, i_2)$ represents the *normalized* joint histogram (*i.e.*, joint pmf) of I_1 and I_2 ('normalized' means that the entries sum up to one).

   For computing the joint histogram, use a bin-width of 10 in both I_1 and I_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. Plot a graph of the values of ρ versus t_x, and another graph of the values of QMI versus t_x.

   Repeat exactly the same steps when the second image is a negative of the first image, i.e. $I_2 = 255 - I_1$.

   Comment on all the plots. In particular, what do you observe regarding the relationship between the dependence measures and the alignment between the two images? Your report should contain all four plots labelled properly, and the comments on them as mentioned before. [25 points]

   Solution: Code in the homework folder. The major point is to realize that when the two images are optimally aligned, the intensities $I_1(x, y)$ and $I_2^{os}(x, y)$ are highly dependent on each other. The superscript *os* stands for 'optimal shift', which in this case is 0. Thus, we would expect the dependence measure to have a maximum for the case when $t_x = 0$. This is obeyed by the QMI but surprisingly not by the correlation coefficient, since the latter does not make use of the full PMF of the intensities (it uses only the mean and variance of the intensities). However if the image intensities obeyed a linear relationship of the form $I_2^{os}(x, y) = \alpha I_1(x, y) + \beta$ where $\alpha \neq 0$, then both QMI and correlation coefficient exhibit the aforementioned property.

   Marking scheme: 5 points for correct joint histogram code, 2 points for computation of marginal using the joint, 3 points for QMI, 1 point for CC (even if inbuilt function is used), 3 points for correct image shifting code (I will allow use of imtranslate from the image processing toolbox). 5 points for plotting the CC and QMI curves in both cases (deduct 3 points if the plots are not included in the report even if the code produces them). 6 points for comments regarding the behaviour of CC and QMI in this application: 2 points for arguing why QMI is superior, 2 points for stating that the dependence measures peak at optimal alignment or near optimal alignment, 2 points for analyzing the case when the two image intensities are linearly related.

7. Derive the covariance matrix of a multinomial distribution using moment generating functions. You are not allowed to use any other method. Since a covariance matrix $\boldsymbol{C}$ is square and symmetric, it is enough to

derive expression for the diagonal elements C_{ii} and the off-diagonal elements $C_{ij}, i \neq j$. [10 points]

Solution: The PMF of a Multinomial distribution with K categories and n trials is given by $\phi_X(\boldsymbol{t}) = (\sum_{i=1}^{K} p_i e^{t_i})^n$ where $\sum_{i=1}^{K} p_i = 1; \forall i 0 \leq p_i \leq 1$, and where $\boldsymbol{t} = (t_1, t_2, ..., t_K)$. Taking partial derivatives w.r.t. t_k, we have $\dfrac{\partial \phi_X(\boldsymbol{t})}{\partial t_k} = n(\sum_{i=1}^{K} p_i e^{t_i})^{n-1} p_k e^{t_k}$. Setting, $\boldsymbol{t} = \boldsymbol{0}$, we have $\dfrac{\partial \phi_X(\boldsymbol{t})}{\partial t_k}|_{\boldsymbol{t}=\boldsymbol{0}} = np_k$. Hence we also have $\dfrac{\partial \phi_X(\boldsymbol{t})}{\partial \boldsymbol{t}}|_{\boldsymbol{t}=\boldsymbol{0}} = n(p_1, p_2, ..., p_k)$.

Note that $C_{kl} = E[(X_k - \mu_k)(X_l - \mu_l)] = E(X_k X_l) - \mu_k \mu_l$ where $\mu_k = E(X_k)$. Now to derive $E(X_k X_l)$, we consider $\dfrac{\partial^2 \phi_X(\boldsymbol{t})}{\partial t_k \partial t_l} = np_k e^{t_k}(n-1)(\sum_{k=1}^{K} p_k e^{t_k})^{n-2} p_l e^{t_l}$. Hence we have $\dfrac{\partial^2 \phi_X(\boldsymbol{t})}{\partial t_k \partial t_l}|_{\boldsymbol{t}=\boldsymbol{0}} = n(n-1)p_k p_l$, which is equal to $E(X_k X_l)$. Hence we get $C_{kl} = n(n-1)p_k p_l - n^2 p_k p_l = -np_k p_l$.

On the other hand, $C_{kk} = E[(X_k - \mu_k)^2] = E(X_k^2) - \mu_k^2$. To obtain $E(X_k^2)$, we obtain $\dfrac{\partial^2 \phi_X(\boldsymbol{t})}{\partial t_k^2} = n(n-1)(\sum_i p_i e^{t_i})^{n-2} p_k^2 e^{t_k} + n(\sum_i p_i e^{t_i})^{n-1} p_k e^{t_k}$. When $\boldsymbol{t} = \boldsymbol{0}$, we get $\dfrac{\partial^2 \phi_X(\boldsymbol{t})}{\partial t_k^2} = n(n-1)p_k^2 + np_k$. Hence $C_{kk} = n(n-1)p_k^2 + np_k - n^2 p_k^2 = np_k(1-p_k)$.

Marking scheme: 5 points for C_{kl} and 5 points for C_{kk}. No credit if MGFs were not used even if the answer is right. For deriving $E(X_k)$ however, the student may use MGFs or any other method. But for $E(X_k X_l)$ or $E(X_k^2)$, only MGFs must be used.