**ISyE 6416: Computational Statistics**
**Homework 1**
**(100 points total.)**

This homework is due on Wed. Jan. 23, 2019, 11:59pm, on canvas. Late submission will not be accepted. Please do not turn in any papers to my mailbox.

- Please remember to staple if you turn in more than one page.
- Please write your team member's name is you collaborate.

1. **Algorithms**

   (a) **Simple questions (10 pts/2.5 pts each question.)**

   - What does algorithm efficiency mean? What are two types of algorithm efficiency measures?
   - What does algorithm robustness mean? Given one example of robust algorithm.
   - What does algorithm stability mean? What's the difference of algorithm stability and robustness?
   - Given two commonly seen definition of algorithm accuracy. Why do we sometimes prefer approximate algorithms?

   (b) **Bisection (20 pts).** Write MATLAB code to implement bisection algorithm to find the $\alpha = 0.95$-quantile of a $t$ distribution with $n = 5$ degrees of freedom. Start with initial interval $[1.291, 2.582]$. Stop when the length of the interval is less than $10^{-4}$. (Hint: You may use for loop in matlab. You do not have to be concerned with efficiency of the code for now.)

   (c) **Worst-case complexity of quicksort (5 pts).** Show that the worst case of quick sort takes $\mathcal{O}(n^2)$ operations.

   (d) **Fourier transform of a delayed signal (5 pts).** Show that

   $$\mathcal{F}(x(t-\tau)) = e^{-i2\pi f\tau}X(f).$$

   (e) **Steps for deriving FFT (20 pts).** Let $x_n$ be a signal that is 0 outside the interval $0 \le n \le N-1$. Suppose $N$ is even. Let $e_n = x_{2n}$ represent the even-indexed samples, and let $o_n = x_{2n+1}$ represent the odd-indexed samples

   i. Show that $e_n$ and $o_n$ are zero outside the interval $0 \le n \le (N/2) - 1$.

   ii. Show that

   $$\widetilde{x}_k = \frac{1}{2}\widetilde{E}_k + \frac{1}{2}W_N^k\widetilde{O}_k, \quad k = 0, 1, \ldots, N-1,$$

   where $W_N = e^{-i\frac{2\pi}{N}}$, and

   $$\widetilde{E}_k = 2\sum_{n=0}^{N/2-1} e_n W_{N/2}^{nk}, \quad \widetilde{O}_k = 2\sum_{n=0}^{N/2-1} o_n W_{N/2}^{nk}.$$

iii. Show that

$$\widetilde{E}_{k+N/2} = \widetilde{E}_k, \quad \widetilde{O}_{k+N/2} = \widetilde{O}_k.$$

2. **Basic linear algebra and statistical inference**

(a) **Rank of a product (5 pts).** Suppose that $A \in \mathbb{R}^{4\times3}$ has rank 2, and $B \in \mathbb{R}^{3\times5}$ has rank 3. What values can $r = \text{Rank}(AB)$ possibly have? For each value $r$ that is possible, given an example, i.e., a specific $A$ and $B$ with the dimensions and ranks given above, for which $\text{Rank}(AB) = r$.

(Optional): (a) Repeat the above questions for $A \in \mathbb{R}^{4\times3}$, $\text{rank}(A) = 2$, $B \in \mathbb{R}^{3\times5}$, $\text{rank}(B) = 1$. (b) Repeat the above questions for $A \in \mathbb{R}^{4\times2}$, $\text{rank}(A) = 2$, $B \in \mathbb{R}^{2\times5}$, $\text{rank}(B) = 1$.

(b) **Simple Bayesian inference (20 pts).**

i. Let $x \sim \mathcal{N}(\mu, \sigma^2)$, and assume a prior distribution $\mu \sim \mathcal{N}(\theta, \tau^2)$. Derive to show that the posterior distribution $\mu|x \sim \mathcal{N}(\frac{\tau^2}{\tau^2+\sigma^2}x + \frac{\sigma^2}{\sigma^2+\tau^2}\theta, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2})$.

ii. Now if we change the prior distribution to be $\mu \sim \text{Unif}[0,1]$, what will be the form of the posterior distribution $\mu|x$?

(c) **Maximum likelihood estimator (10 pts).** Let $X_1, \ldots, X_n$ independent random variables identically distributed with density function

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

- Find the Maximum Likelihood Estimator for $a \in \mathbb{R}$ and $b$.
- (Optional) Are they unbiased estimators? (Hint: consider order statistics).

(d) **Hypothesis test of the mean (5 pts).** The drying time for a certain type of paint under specified test conditions is known to be normally distributed with mean value 75 min and standard deviation $\sigma = 9$ min. Chemists have proposed a new additive designed to *decrease* average drying time. It is believed that the drying times with this additive will remain normally distributed with $\sigma = 9$. Because of the expense associated with the additive, evidence should strongly suggest an improvement in average drying time before such a conclusion is adopted. Experimental data consist of drying times from $n = 25$ test specimens. State the hypotheses to be tested. Construct a likelihood ratio test and calculate the threshold so that the probability of false detection is 0.05.