ISyE 6416: Computational Statistics **(Due: 01/22/2020)**

Homework Assignment #1

Instructor: Xie, Yao *Name:* Chen-Yang(Jim), Liu *GTID:* 90345****

Problem 1: Algorithms (10+20+5+5+20=60 points)

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

- What does algorithm efficiency mean? What are two types of algorithm efficiency measures?
  1. **Algorithm Efficiency:**
     A usual method to appraise the quality of algorithm efficiency is speed. That is, we would like to know how long an algorithm runs to produce its result.
  2. **Two Types:**
     (a) Space Efficiency
     the memory we need to store an algorithm
     (b) Time Efficiency
     the time required to go through a series of steps of an algorithm
- What does algorithm robustness mean? Given one example of robust algorithm.
  A robust algorithm can be applied to a wide range of data. An example of this is robust optimization.
  $$\min_x c^T x : a_i \in \mu_i, a_i^T \leq b_i, i = 1, ..., m$$
  This is a Robust Linear Programming which can be used when data is uncertain and a solution still can be derived.
- What does algorithm stability mean? What's the difference of algorithm stability and robustness?
  Stable algorithm means that a little perturbation does not affect big difference in deriving solutions. The key difference between robustness and stability is that robust program will detect inappropriate data input for both the algorithm and the implementation of the program. On the other hand, stable algorithms run appropriate data input and then a small difference in data input does not create largely different solutions.
- Given two commonly seen definition of algorithm accuracy. Why do we sometimes prefer approximate algorithms?
  Although some algorithms are exact, other algorithms are usually approximate since some of computed results does not have specific closed-form solution. For example, Taylor expansion says a function could be approximated by multiple derivatives of functions. That is an approximate algorithm.

(b) Bisection

```
from scipy.stats import t
import matplotlib.pyplot as plt
def bisection(f,a,b):
    '''Approximate solution of f(x)=0 on interval [a,b] by bisection method.
    ----------
    f : function
        The function for which we are trying to approximate a solution f(x)=0.
    a,b : numbers
        The interval in which to search for a solution. The function returns
        None if f(a)*f(b) >= 0 since a solution is not guaranteed.
```

```
    Returns
    -------
    x_N : number
        The midpoint of the Nth interval computed by the bisection method. The
        initial interval [a_0,b_0] is given by [a,b]. If f(m_n) == 0 for some
        midpoint m_n = (a_n + b_n)/2, then the function returns this solution.
        If all signs of values f(a_n), f(b_n) and f(m_n) are the same at any
        iteration, the bisection method fails and return None.
    --------
    '''
    if f(a)*f(b) >= 0:
        print("Bisection method fails.")
        return None
    a_n = a
    b_n = b
    iteration = 0
    while (b - a > (10 ** -4)):
        m_n = (a_n + b_n)/2
        f_m_n = f(m_n)
        if f(a_n)*f_m_n < 0:
            a_n = a_n
            b_n = m_n
        elif f(b_n)*f_m_n < 0:
            a_n = m_n
            b_n = b_n
        elif f_m_n == 0:
            print("Found exact solution.")
            print("Iteration times: ", iteration)
            print("Interval is: ", a_n, b_n)
            return 'Solution point is: ', m_n
        else:
            print("Bisection method fails.")
            print("Iteration times: ", iteration)
            return None
        iteration += 1
    return (a_n + b_n)/2

f = lambda x: t.cdf(x, df = 5, loc=0, scale=1) - 0.95
bisection(f, 1.291,  2.582)

Found exact solution.
('Iteration times: ', 48)
('Interval is: ', 2.0150483733330207, 2.0150483733330256)
```

(c) Worst-case complexity of quicksort

Worst case means: Given a strictly decreasing sequence with n numbers, i.e. $e_i < e_j\ for\ i\ <\ j\ where\ i,\ j\ = 1,\ 2,\ ....,\ n$. Now pick e_n as a pivot and from the following calculation, $O(n^2)$ is time complexity of the worst case.

$$\begin{aligned}
e_n &= n - 1 + \sum_{i=1}^{n-1} e_i \\
&= n - 1 + n - 2 + \sum_{i=2}^{n-1} e_i \\
&= n - 1 + n - 2 + ... + 1 \\
&= \frac{n(n-1)}{2} = \frac{n^2 - n}{2}
\end{aligned}$$

(d) Fourier transform of a delayed signal

$$\begin{aligned}
F(x(t-\tau)) &= F(x(t-\tau) * h) \ where \ h = 1 \\
&= \iint x(t-\tau)e^{-i2\pi ft} d\tau dt \\
&= \int [\int x(t-\tau)e^{-i2\pi ft} dt] d\tau \\
&= X(f) \int e^{-i2\pi f\tau} * 1 d\tau \\
&= X(f)e^{-i2\pi f\tau} \quad where \int e^{-i2\pi f\tau} * 1 d\tau = 1
\end{aligned}$$

(d) Fourier transform of a delayed signal

$$\begin{aligned}
F(x(t-\tau)) &= F(x(t-\tau) * h) \ where \ h = 1 \\
&= \iint x(t-\tau)e^{-i2\pi ft} d\tau dt \\
&= \int [\int x(t-\tau)e^{-i2\pi ft} dt] d\tau \\
&= X(f) \int e^{-i2\pi f\tau} * 1 d\tau \\
&= X(f)e^{-i2\pi f\tau} \quad where \int e^{-i2\pi f\tau} * 1 d\tau = 1
\end{aligned}$$

(e) Steps for deriving FFT

(i) $e_n : x_{2n}$ is even-indexed samples and $o_n : x_{2n+1}$ is odd-indexed samples Given $x_n = 0$ if $n \notin [0, N-1]$ and N is even Shift coverage, i.e. $n \notin [1, N]$ Since $e_n + o_n = x_{2n} + x_{2n+1} = 0$ i.e. x_{2n} where N/2 samples are not equal to 0, $x_{2n} = 0 \ if \ n \notin [1, N/2]$ Then, we shifted coverage again. We get $n \notin [0, N/2-1]$

(ii) Goal: $\tilde{x} = \sum_{j=0}^{N-1} x_j e^{-i\frac{2\pi}{N}jk} = \sum_{j=0}^{N-1} x_j W_N$

$$\begin{aligned}
\tilde{x} &= \sum_{n=0}^{N-1} x_n e^{-i\frac{2\pi}{N}nk} \\
&= \sum_{n=0,even}^{N/2-1} x_n e^{-i\frac{2\pi}{N}nk} + \sum_{n=0,odd}^{N/2-1} x_n e^{-i\frac{2\pi}{N}nk} \\
&= \sum_{n=0}^{N/2-1} x_{2n} e^{-i\frac{2\pi}{N}2nk} + \sum_{n=0}^{N/2-1} x_{2n+1} e^{-i\frac{2\pi}{N}(2n+1)k} \\
&= \sum_{n=0}^{N/2-1} e_n e^{-i\frac{2\pi}{N/2}nk} + \sum_{n=0}^{N/2-1} o(n) e^{-i\frac{2\pi}{N/2}nk} e^{-i\frac{2\pi}{N}k} \\
&= \frac{1}{2} \sum_{n=0}^{N/2-1} 2e_n (W_{N/2})^{kn} + \frac{1}{2} W_n \sum_{n=0}^{N/2-1} 2o_n (W_{N/2})^{kn} \\
&= \frac{1}{2}\tilde{E_k} + \frac{1}{2} W_N^k \tilde{O_k}
\end{aligned}$$

(iii)

$$
\begin{aligned}
\tilde{E}_{k+N/2} &= 2\sum_{n=0}^{N/2-1} e_n W_{N/2}^{n(k+N/2)} \\
&= 2\sum_{n=0}^{N/2-1} e_n W_{N/2}^{nk} W_{N/2}^{\frac{nN}{2}} \\
&= 2\sum_{n=0}^{\frac{N}{2}-1} e_n e^{-i\frac{2\pi}{\frac{N}{2}}nk} e^{-i\frac{2\pi}{\frac{N}{2}}\frac{nN}{2}}
\end{aligned}
$$

Now consider $e^{-i\frac{2\pi}{\frac{N}{2}}\frac{nN}{2}}$ and by Euler formula,

$$
\begin{aligned}
e^{-i\frac{2\pi}{\frac{N}{2}}\frac{nN}{2}} &= e^{-i2n\pi} \\
&= \cos(2n\pi) - isin(2n\pi) \\
&= 1 - 0 = 1 \qquad (1)
\end{aligned}
$$

By (1),

$$
2\sum_{n=0}^{\frac{N}{2}-1} e_n e^{-i\frac{2\pi}{\frac{N}{2}}nk} e^{-i\frac{2\pi}{\frac{N}{2}}\frac{nN}{2}} = 2\sum_{n=0}^{\frac{N}{2}-1} e_n e^{-i\frac{2\pi}{\frac{N}{2}}nk} = \tilde{E}_k
$$

Similarly,

$$
\begin{aligned}
\tilde{O}_{k+N/2} &= 2\sum_{n=0}^{\frac{N}{2}-1} o_n e^{-i\frac{2\pi}{\frac{N}{2}}nk} \\
&= \tilde{O}_k
\end{aligned}
$$

Problem 2: Basic linear algebra and statistical inference (5+20+10+5=40 points)

(a) Rank of a product

$$A_{4\times 3} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix}, B_{3\times 5} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \end{bmatrix} AB_{4\times 5} = \begin{bmatrix} Ab_1 & Ab_2 & Ab_3 & Ab_4 & Ab_5 \end{bmatrix}$$

Now we could investigate some cases about AB matrix from the given assumption,

- B is singular B is singular $\Rightarrow$ rank(AB)$\leq$ rank(B)
- A is singular A is singular $\Rightarrow$rank(AB) $\leq$ rank(A)

Therefore, rank(AB) depends on whether A or B is singular. i.e. rank(AB) $\leq$ min(rank(A), rank(B)) Back to question, value r is 2.

(b) Simple Bayesian inference

1. Prior:$p(\mu) \sim N(\theta, \tau^2)$, Observed data $x \sim N(\mu, \sigma^2)$
   Those two distributions are i.i.d. And posterior distribtion is:

$$\begin{aligned}
f(\mu \mid x) &= \frac{f(x \mid \mu)p(\mu \mid \tau)}{f(\mu)} \\
f(\mu \mid x) &= \Pi_{i=1}^{n=1} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}} \times \frac{1}{\sqrt{2\pi}\tau} e^{\frac{-(\mu-\theta)^2}{2\tau^2}} \\
&\propto e^{-\frac{1}{2}\sum_{i=1}^{n=1} \frac{(x_i-\mu)^2}{\sigma^2} + \frac{(\mu-\theta)^2}{\tau^2}} \\
&= e^{-\frac{1}{2}\frac{(\sum_{i=1}^{n=1} x_i^2 - 2\mu\sum_{i=1}^{n=1} x_i + n\mu^2)}{\sigma^2} + \frac{\mu^2-2\mu\theta+\theta^2}{\tau^2}} \\
&= e^{-\frac{1}{2}[(\frac{n}{\sigma^2}+\frac{1}{\tau^2})\mu^2 - 2(\frac{n\bar{x}}{\sigma^2}+\frac{\theta}{\tau^2})\mu + \frac{(\sum_{i=0}^{n=1} x_i^2)}{\sigma^2} + \frac{\theta^2}{\tau^2}]} \\
&\propto e^{-\frac{1}{2}\frac{(\mu-\mu_{posterior})^2}{\tau^2_{posterior}}} \\
&= e^{-\frac{1}{2}\frac{\mu^2-2\mu\mu_{posterior}+\mu^2_{posterior}}{\tau^2_{posterior}}} \quad where\ n = 1\ i.e. x = x_1 \\
\frac{1}{\tau^2_{posterior}} &= \frac{1}{\sigma^2} + \frac{1}{\tau^2} \implies \tau_{posterior} = \frac{\tau^2\sigma^2}{\tau^2+\sigma^2} \\
\frac{\mu_{posterior}}{\tau^2_{posterior}} &= \frac{x}{\sigma^2} + \frac{\theta}{\tau^2} \implies \mu_{posterior} = \frac{\tau^2}{\tau^2+\sigma^2}x + \frac{\sigma^2}{\tau^2+\sigma^2}\theta \\
&\sim N(\frac{\tau^2}{\tau^2+\sigma^2}x + \frac{\sigma^2}{\tau^2+\sigma^2}\theta, \frac{\tau^2\sigma^2}{\tau^2+\sigma^2})
\end{aligned}$$

2. Prior: $p(\mu) \sim Uniform(0, 1)$, observed data $x \sim N(\mu, \sigma^2)$
   Those two distributions are i.i.d. And posterior distribtion is:

$$\begin{aligned}
f(\mu \mid x) &= \frac{f(x \mid \mu)p(\mu \mid \tau)}{f(\mu)} \\
f(\mu \mid x) &= \Pi_{i=1}^{n=1} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}} \times 1 \\
&\sim N(\mu, \sigma^2)
\end{aligned}$$

(c) Maximum likelihood estimator

(i) Find MLE

$$
\begin{aligned}
L(a,b \mid x) &\triangleq \Pi_{i=1}^{n} f(x) = (\frac{1}{b-a})^n I(a,b)^n \\
&= (\frac{1}{b-a})^n I(a \le x_i) I(x_i \le b)
\end{aligned}
$$

According to MLE, our goal is to find a parameter that most likely occurrs.
Now, we have some cases to consider:

$$\hat{a} = arg \max_{a} L(a,b \mid x) \implies \hat{a} = X_{(1)} \; [order\ statistics, b\ fixed]$$

$$\hat{b} = arg \max_{b} L(a,b \mid x) \implies \hat{b} = X_{(n)} [order\ statistics, a\ fixed]$$

(ii) Are $\hat{a}, \hat{b}$ unbiased estimators?

(d) Hypothesis test of the mean

Hypothese Testing:

$$H_0 : \mu = \mu_0 = 75 \;\; [Null\ Hypothesis], H_1 : \mu = \mu_1 < \mu_0 \;\; [Alternative\ Hypothesis]$$

In this case, we try to know additive is effective or not. H_0 means new additive does not have any effect.
Likelihood Ration Test:

$$
\begin{aligned}
\lambda(\mu_0, \mu_1) &= \Pi_{i=1}^{25} \frac{1}{\sqrt{2\pi 9}} e^{\frac{-(x_i-\mu_1)^2}{2\times 9^2}} / \Pi_{i=1}^{25} \frac{1}{\sqrt{2\pi 9}} e^{\frac{-(x_i-75)^2}{2\times 9^2}} \\
&= e^{(-\sum_{i=1}^{25}(x_i-\mu_1)+\sum_{i=1}^{25}(x_i-75)^2)/2\times 9^2} > b \\
\ell(\lambda(\mu_0,\mu_1)) &= -\sum_{i=1}^{25}(x_i-\mu_1) + \sum_{i=1}^{25}(x_i-75)^2 > b \times 2 \times 9^2 = b_1 \\
&\implies \sum_{i=1}^{25} -2\mu_1 x_i + \sum_{i=1}^{25} -150x_i + 75^2 - \mu_1^2 > b1 \\
&\implies \sum_{i=1}^{25} x_i(-2\mu_i - 150) > b1 - 75^2 + \mu_1^2 = b_2 \; where \; -2\mu_i - 150 < 0 \\
&\implies (\sum_{i=1}^{25} x_i)/25 < b_2(-2\mu_i - 150)/25 = b_3 \\
&s.t.\ P(\bar{x} < b_3 \mid \mu = \mu_0) = \alpha = 0.05 \\
&i.e.\ P(\bar{x} - \mu_0/\sqrt{18/25} < -Z_{0.05,n=25} \mid \mu = \mu_0) = 0.05 \\
&i.e.\ P(\bar{x} - \mu_0/\sqrt{18/25} < -Z_{0.05,n=25} \mid \mu = \mu_0) = 0.05 \\
&i.e.\ P(\bar{x} - \mu_0/\sqrt{18/25} < -1.645 \mid \mu = \mu_0) = 0.05
\end{aligned}
$$

From the above formula, the threshold is $-1.645 \times \sqrt{18/25} + 75 = -1.645 \times 0.8485281374 + 75 = 73.60417121$.