IE531: Algorithms for Data Analytics
Midterm Exam
Spring, 2020
5:00PM-7:00PM, March 29, 2020
Zoom URL:
https://illinois.zoom.us/j/7779026329

Instructions

1. **A (correct) answer that is not accompanied by an appropriate justification will not be given any credit. Make sure you answer all questions.**
2. **Try to finish on time.**

Instructions

1. (*10 points*) **Recursion/Repeated-Squaring-Algorithm**: Consider the pseudo-code shown in figures 1(a) and 1(b). Both these code samples compute the value for x^n for any $x \in \mathcal{R}$ and $n \in \mathcal{N}$. What is the running-time of each of these implementations? In two/three sentences can you tell me which one of them would be faster for the same inputs? Give me a clear reason to receive full-credit for this problem.

```
double power1(double x, int n)
{
        if (n == 0) return 1.0;
        return x*power1(x,n-1);
}
```

(a) Problem 1(a)

```
double power2(double x, int n)
{
        if (n == 0)
                return 1.0;
        double p = power2(x,n/2);
        if (n%2 == 0)
                return p*p;
        else
                return x*p*p;
}
```

(b) Problem 1(b)

Figure 1: (Two versions of the) Code for problem 1.

Both these code samples compute the value of x^n where x (n) is a double (int). Come see me if you are lost – figure 1(b) is known as the method of *repeated squaring*. The method in figure 1(b) is faster than the one in figure 1(a). The appropriate version of Theorem 3.1 of Lesson 1 would be

$$\begin{aligned}\text{power1} : T(n) &= 1 \times T(n-1) + c \\ \text{power2} : T(n) &= 1 \times T(n/2) + c.\end{aligned}$$

power1 will take $O(c \times n)$ time to complete, while power2 will take $O(log_2 n)$ time to complete (why? $a = 1, b = 2, k = 0$ for this case). power2 is obviously faster.

Partial Credit No scope.

2. (*10 points*) **Transforming Discrete RVs to Continuous RVs**: In Programming Assignment #4 you considered a 3-sided dice as shown in figure 2 that is not necessarily fair. With this 3-sided dice, you have i.i.d. discrete random variables $\{x_i\}_{i=1}^{\infty}$, where $x_i \in \{1, 2, 3\}$ where the probabilities of the three-possible events are **unknown**. That is, $Prob(x_i = 1) = p_1, Prob(x_i = 2) = p_2$, and $Prob(x_i = 3) = (1 - p_1 - p_2)$, where p_1 and p_2 are not known. How would you generate a sequence of i.i.d. unit-normal, random variables $\{y_i\}_{i=1}^{\infty}$ using **just** this unknown distribution?

Figure 2: Three-sided-dice.

There are many solutions to this problem – the meta-principle (following that of the solution to the previous problem) is to find two compound-events with equal probability that we interpret as "fair-heads" and "fair-tails;" following this, we just use the procedure of Problem 4 of Quiz 8 to create uniform i.i.d. r.v.s.

Approach 1: Borrowing from the solution to the previous problem, we make two calls to the i.i.d. discrete r.v. generator. If this results in $\langle 1, 2\rangle$ we interpret it as a "fair-heads" and if we get $\langle 2, 1\rangle$ we call it a "fair-tail." It is not hard to see that $Prob(\langle 1, 2\rangle) = Prob(\langle 1, 2\rangle) = p_1 p_2$. We reject all other outcomes (like we rejected TT and HH in the solutions to the previous problem). The fair-coin can be used to create an u.i.i.d. r.v. and then use the Box-Muller method to generate a unit normal r.v.

Approach 2: The previous approach will reject a large-number of outcomes (leading to inefficiency). If we looked at all possible pairs of events, we get their probabilities to be

$$
\begin{array}{lll}
Prob(\langle 1,1\rangle) = p_1^2 & Prob(\langle 2,1\rangle) = p_1 p_2 & Prob(\langle 3,1\rangle) = p_1 p_3 \\
Prob(\langle 1,2\rangle) = p_1 p_2 & Prob(\langle 2,2\rangle) = p_2^2 & Prob(\langle 3,2\rangle) = p_2 p_3 \\
Prob(\langle 1,3\rangle) = p_1 p_3 & Prob(\langle 2,3\rangle) = p_2 p_3 & Prob(\langle 3,3\rangle) = p_3^2
\end{array}
$$

We reject all event-pairs along the diagonal of the above matrix (i.e. reject $\langle 1, 1\rangle, \langle 2, 2\rangle$ and $\langle 3, 3\rangle$ events). We split the off-diagonal events into two groups with equal probabilities of occurrence as follows ($p_3 = 1 - p_1 - p_2$):

$$\text{“fair-heads”} \equiv \{\langle 1,2\rangle, \langle 1,3\rangle, \langle 2,3\rangle\}$$

$$\text{“fair-tails”} \equiv \{\langle 2,1\rangle, \langle 3,1\rangle, \langle 3,2\rangle\}$$

It is not hard to see that $Prob(\text{“fair-heads”}) = Prob(\text{“fair-tails”}) = p_1p_2 + p_1p_3 + p_2p_3$, while $Prob(\text{reject}) = p_1^2 + p_2^2 + p_3^2$. We then use the method of problem 4 of Quiz 8 to create i.i.d. uniform r.v.s from this fair-coin-toss, and following this we just use the Box-Muller method to generate unit-normal r.v.s.

Grading Observations

(a) Getting at a method that gets a “fair coin from an unfair-coin” gave you 7.5 points.

(b) Transforming the outcome of the resulting fair-coin into u.i.i.d r.v.s and then using the inverse-transform method got you 2.5 points.

3. (*10 points*) **Singular Value Decomposition**: Consider the following SVD of a matrix $\mathbf{A} = \mathbf{UDV}^T$, where

$$\underbrace{\begin{pmatrix} 4.7511 & 1.9078 & 0.9344 & 3.2316 \\ 0.1722 & 3.8276 & 2.4488 & 3.5468 \\ 2.1937 & 3.9760 & 2.2279 & 3.7734 \end{pmatrix}}_{=\mathbf{A}} = \underbrace{\begin{pmatrix} -0.5493 & -0.7744 & -0.3139 \\ -0.5379 & 0.6152 & -0.5763 \\ -0.6394 & 0.1477 & 0.7545 \end{pmatrix}}_{=\mathbf{U}} \underbrace{\begin{pmatrix} 9.8247 & 0 & 0 \\ 0 & 3.7411 & 0 \\ 0 & 0 & 0.2958 \end{pmatrix}}_{=\mathbf{D}}$$
$$\underbrace{\begin{pmatrix} -0.4179 & -0.5750 & -0.3313 & -0.6205 \\ -0.8685 & 0.3915 & 0.2973 & 0.0633 \\ 0.2178 & 0.6594 & -0.0800 & -0.7151 \end{pmatrix}}_{=\mathbf{V}^T}$$

(a) (*5 points*) What is the best-possible rank 1 approximation to $\mathbf{A}$? (**PS**: No need to carry out the multiplication; just give me the form/structure)

$$\mathbf{A}_1 = 9.8247 \times \begin{pmatrix} -0.5493 \\ -0.5379 \\ -0.6394 \end{pmatrix} \times \begin{pmatrix} -0.4179 & -0.5750 & -0.3313 & -0.6205 \end{pmatrix}$$
$$\left(= \mathbf{U} \times \begin{pmatrix} 9.8247 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \times \mathbf{V}^T \right)$$

(b) (*5 points*) What is the best-possible rank 4 approximation to $\mathbf{A}$? (**PS**: No need to carry out the multiplication; just give me the form/structure)

If the “approximation matrix” $\mathbf{B}$ should be a 3×4 matrix, then there can be **no B** with rank 4 (why?).

If $\mathbf{B}$ does not <u>*have*</u> to be a 3×4 matrix, but it <u>*has*</u> to be of rank 4, such that the $\|\mathbf{B}\|_F \approx \|\mathbf{A}\|_F$, then $\mathbf{B}$ is not unique. To see this, note that for any $0 < \epsilon$,

$$\mathbf{B}(\epsilon) = \begin{pmatrix} 4.7511 & 1.9078 & 0.9344 & 3.2316 \\ 0.1722 & 3.8276 & 2.4488 & 3.5468 \\ 2.1937 & 3.9760 & 2.2279 & 3.7734 \\ 0 & 0 & 0 & \epsilon \end{pmatrix}, rank(\mathbf{B}(\epsilon)) = 4, \text{ and } \|\mathbf{B}(\epsilon)\|_F = \|\mathbf{A}\|_F + O(\epsilon^2)$$

Therefore, by making ϵ smaller-and-smaller $\|\mathbf{B}(\epsilon)\|$ will get closer-and-closer to $\|\mathbf{A}\|_F$. But, ϵ cannot be equal to zero (because, if it did, the $rank(\mathbf{B}(\epsilon)) \neq 4$).

Either way, there is no unique best-possible rank 4 approximation to $\mathbf{A}$. This was the subject of the Flipped-Classroom Video with the title *Best-Rank-K-Approximation Problem: The algorithm and some subtle theoretical-issues* on Compass for Lesson 3.

While we are at it, assuming you caught the non-uniqueness argument presented above, you may think the same issue will come to haunt us when we are looking at rank 3, rank 2 or rank 1 approximations of $\mathbf{A}$. It turns out that all the lower-rank approximation $\mathbf{B}$ are unique if the singular values of $\mathbf{A}$ are unique (i.e. there are no repeated singular-values).

4. (*10 points*) **Singular Value Decomposition**: Let $\mathbf{A}$ be a matrix. How would you compute

$$\mathbf{v}_1 = \underset{\|\mathbf{v}_1\| = 1}{arg\ max} \|\mathbf{A}\mathbf{v}_1\|?$$

(**From Section 3.7 of the Book** – You compute $\mathbf{B} = \mathbf{A}^T\mathbf{A}$, which is a square-symmetric matrix, where

$$\mathbf{B} = \sum_{I=1}^{r} \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$$

Use repeated-squaring to compute a "large" power/exponent of $\mathbf{B}$, that is compute $\mathbf{B}^k$ for some large value of k, where

$$\mathbf{B}^k = \sum_{I=1}^{r} \sigma_i^{2k} \mathbf{v}_i \mathbf{v}_i^T$$

If $\sigma_1 > \sigma_2$, then the first-term of the summation dominates the rest, so $\mathbf{B}^k \approx \sigma_1^{2k}\mathbf{v}_1\mathbf{v}_1^T$. This means a close estimate to $\mathbf{v}_1$ can be computed by simply taking the first-column of $\mathbf{B}^k$ and normalizing it to a unit-vector.

There are other methods listed in the book, as well. Any one of them would have gotten you full-points. No partial credit.

5. (*20 points*) **Markov Chains, Stationary Probability Distributions, etc**: Consider a Markov Chain with a probability matrix as shown below

$$\mathbf{P}_1 = \begin{pmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} & 0 \end{pmatrix}$$

(a) (*2.5 points*) Is this Markov Chain *irreducible*?

From the graphical description of the Markov Chain that represents the stochastic-matrix $\mathbf{P}_1$(cf. figure 3) we see that we move from any state to any other state – that is, the Markov Chain is *strongly-connected*, therefore the Markov Chain represented by $\mathbf{P}_1$ is irreducible.

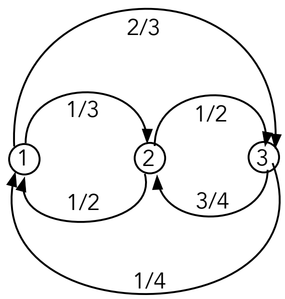

Figure 3: Markov Chain that represents the stochastic-matrix $\mathbf{P}_1$ of problem 6a.

(b) (*2. 5 points*) Is this Markov Chain *aperiodic*?

From figure 3 we see that the graphical description of the Markov Chain has loops of lengths 2 and 3. Since $gcd(2,3) = 1$, this Markov Chain is *aperiodic*.

(c) (*2.5 points*) Show that this Markov Chain has a unique stationary probability distribution.

From Theorem 4.2 of the Book, every irreducible (i.e. strongly connected) Markov Chain has a unique stationary probability distribution. Since this Markov Chain is irreducible (cf. solution to problem 5a), it follows that there is a unique stationary probability distribution for this Markov Chain.

(d) (*5 points*) What is the stationary probability distribution of this Markov Chain?

There are many ways to getting at the (unique) stationary probability distribution of this Markov Chain. I am going to use the method in the proof of Theorem 4.2 in the book.

$$\mathbf{A} = \begin{pmatrix} (\mathbf{P}_1 - \mathbf{I}) & \mathbf{1} \end{pmatrix} = \begin{pmatrix} -1 & \frac{1}{3} & \frac{2}{3} & 1 \\ \frac{1}{2} & -1 & \frac{1}{3} & 1 \\ \frac{1}{4} & \frac{3}{4} & -1 & 1 \end{pmatrix} \Rightarrow \mathbf{B} = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 1 \\ -1 & \frac{1}{3} & 1 \\ \frac{3}{4} & -1 & 1 \end{pmatrix}$$

$$\Rightarrow \pi = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 1 \\ -1 & \frac{1}{3} & 1 \\ \frac{3}{4} & -1 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 0.2727 & 0.3636 & 0.3636 \end{pmatrix}$$

(e) (*2.5 points*) Certify/verify your answer to problem 5d

$$\underbrace{\begin{pmatrix} 0.2727 & 0.3636 & 0.3636 \end{pmatrix}}_{\boldsymbol{\pi}} \times \underbrace{\begin{pmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} & 0 \end{pmatrix}}_{\mathbf{P}_1} = \underbrace{\begin{pmatrix} 0.2727 & 0.3636 & 0.3636 \end{pmatrix}}_{\boldsymbol{\pi}}$$

(f) (*5 points*) What is $\lim_{n\to\infty} \mathbf{P}_1^n$?

Since the Markov Chain is irreducible (cf. solution to problem 5b), it follows that

$$\lim_{n\to\infty} \mathbf{P}_1^n = \begin{pmatrix} \boldsymbol{\pi} \\ \cdots \\ \boldsymbol{\pi} \end{pmatrix}$$

That is, the resulting matrix will have identical rows, and each row will be $\boldsymbol{\pi}$.

6. (*20 points*) **Markov Chains, Stationary Probability Distributions, etc**: Consider a Markov Chain with a probability matrix as shown below

$$\mathbf{P}_2 = \begin{pmatrix} 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

(a) (*2.5 points*) Is this Markov Chain *irreducible*?

From the graphical description of the Markov Chain that represents the stochastic-matrix $\mathbf{P}_2$(cf. figure 4) we see that we move from any state to any other state – that is, the Markov Chain is *strongly-connected*, therefore the Markov Chain represented by $\mathbf{P}_2$ is irreducible.

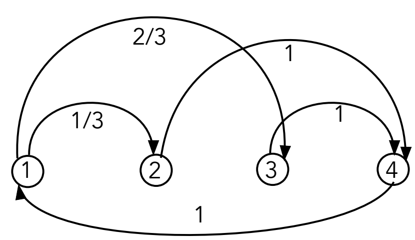

Figure 4: Markov Chain that represents the stochastic-matrix $\mathbf{P}_2$ of problem 6.

(b) (*2. 5 points*) Is this Markov Chain *aperiodic*?

All loops in the graphical representation of the Markov Chain represented by the stochastic-matrix $\mathbf{P}_2$ have a length of 3. Therefore, this Markov Chain is <u>*not*</u> aperiodic.

(c) (*2.5 points*) Show that this Markov Chain has a unique stationary probability distribution.

From Theorem 4.2 of the Book, every irreducible (i.e. strongly connected) Markov Chain has a unique stationary probability distribution. Since this Markov Chain is irreducible (cf. solution to problem 6a), it follows that there is a unique stationary probability distribution for this Markov Chain.

(d) (*5 points*) What is the stationary probability distribution of this Markov Chain?

There are many ways to getting at the (unique) stationary probability distribution of this Markov Chain. I am going to use the method in the proof of Theorem 4.2 in the book.

$$\mathbf{A} = \begin{pmatrix} (\mathbf{P}_2 - \mathbf{I}) & \mathbf{1} \end{pmatrix} = \begin{pmatrix} -1 & \frac{1}{3} & \frac{2}{3} & 0 & 1 \\ 0 & -1 & 0 & 1 & 1 \\ 0 & 0 & -1 & 1 & 1 \\ 1 & 0 & 0 & -1 & 1 \end{pmatrix} \Rightarrow \mathbf{B} = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 1 \\ -1 & 0 & 1 & 1 \\ 0 & -1 & 1 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

$$\Rightarrow \boldsymbol{\pi} = \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 1 \\ -1 & 0 & 1 & 1 \\ 0 & -1 & 1 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 0.3333 & 0.1111 & 0.2222 & 0.3333 \end{pmatrix}$$

(e) (*2.5 points*) Certify/verify your answer to problem 6d

$$\underbrace{\begin{pmatrix} 0.3333 & 0.1111 & 0.2222 & 0.3333 \end{pmatrix}}_{\boldsymbol{\pi}} \times \underbrace{\begin{pmatrix} 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}}_{\mathbf{P}_2} = \underbrace{\begin{pmatrix} 0.3333 & 0.1111 & 0.2222 & 0.3333 \end{pmatrix}}_{\boldsymbol{\pi}}$$

(f) (*5 points*) What is $\lim_{n\to\infty} \mathbf{P}_2^n$?

The Markov Chain represented by the stochastic-matrix $\mathbf{P}_2$ is not aperiodic. Therefore, the above limit is not defined. That said, it follows that (cf. Theorem 4.2 of the book) that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbf{P}_2^i = \begin{pmatrix} \boldsymbol{\pi} \\ \cdots \\ \boldsymbol{pi} \end{pmatrix}$$

7. (*10 points*) **MCMC-MH**: Using the unfair 3-sided dice of figure 2 as your only source of randomness (i.e. this is your *proposal distribution*) construct the Probability Matrix of a 5-state Markov Chain whose stationary probability distribution is $(\frac{1}{5}\ \frac{1}{5}\ \frac{1}{5}\ \frac{1}{5}\ \frac{1}{5})$.

I am going to show how the unfair 3-sided dice of figure 2 can be used to construct (a stochastic-matrix that corresponds to) a Markov Chain on n-many states. You can make $n = 5$, and get the solution to the above problem.

$$\mathbf{P} = \begin{pmatrix} p_1 & p_2 & p_3 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & p_1 & p_2 & p_3 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & p_1 & p_2 & p_3 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & p_1 & p_2 & p_3 \\ p_3 & 0 & 0 & 0 & 0 & \cdots & 0 & p_1 & p_2 \\ p_2 & p_3 & 0 & 0 & 0 & \cdots & 0 & 0 & p_1 \end{pmatrix},$$

where $p_3 = 1 - p_1 - p_2$. It is not hard to see that each *column* of the above matrix adds up to 1 (this is in addition the fact that each *row* adds up to 1). That is, the above $n \times n$ matrix is a doubly-stochastic-matrix, and it is not hard to show that

$$\begin{pmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix} \times \begin{pmatrix} p_1 & p_2 & p_3 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & p_1 & p_2 & p_3 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & p_1 & p_2 & p_3 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & p_1 & p_2 & p_3 \\ p_3 & 0 & 0 & 0 & 0 & \cdots & 0 & p_1 & p_2 \\ p_2 & p_3 & 0 & 0 & 0 & \cdots & 0 & 0 & p_1 \end{pmatrix} = \begin{pmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$$

8. (*10 points*) **MCMC-MH**: You are a store inspector for *Walmart*, your job involves visiting various stores in the Midwest to inspect their operations. You have access to the net-sales figures of all stores within your jurisdiction. Your objective is to pay "surprise-visits" to each store, and the number of times you visit a store should be proportional to its net-sales (i.e. if a store sells more product, you have make more "surprise visits" to that store). Assume you have access to a uniform RV generator, describe a provably-correct algorithm that schedules these "surprise visits."

   The probability that you visit the i-th store is

   $$\pi_i = \frac{\text{sales at the } i\text{-th store}}{\text{total-sales at all stores}}$$

   You can pick a value for r that is less-than-or-equal-to the total number of stores in the Midwest. You then use the MCMC-MH algorithm covered in class to generate a stochastic-matrix that has the desired stationary-distribution. At the end of each visit, you pick the next-store to visit randomly using the stochastic-matrix that you got. The result will be a random (i.e. unpredictable) inspection-sequence that will make you visit the stores as per their sales-figures.