EECS 126 Probability and Random Processes University of California, Berkeley: Spring 2016
Kannan Ramchandran February 16, 2016

Midterm Exam 1 (**Solutions**)

Last name	First name	SID

Name of student on your left:
Name of student on your right:

- DO NOT open the exam until instructed to do so.
- Note that the test has **104** points. but a score ≥ 100 is considered perfect.
- You have 10 minutes to read this exam without writing anything and 90 minutes to work on the problems.
- Box your final answers.
- **Remember to write your name and SID on the top left corner of every sheet of paper.**
- **Do not write on the reverse sides of the pages.**
- All electronic devices must be turned off. Textbooks, computers, calculators, etc. are prohibited.
- No form of collaboration between students is allowed. If you are caught cheating, you may fail the course and face disciplinary consequences.
- **You must include explanations to receive credit.**

Problem	Part	Max	Points	Problem	Part	Max	Points
1	(a)	8		2		20	
	(b)	6		3		15	
	(c)	6		4		12	
	(d)	6		5		25	
	(e)	6					
		32					
Total						104	

Cheat sheet 1. Discrete Random Variables

1) Geometric with parameter $p \in [0,1]$:
$P(X=n) = (1-p)^{n-1}p,\ n \geq 1$
$E[X] = 1/p,\ \text{var}(X) = (1-p)p^{-2}$

2) Binomial with parameters N and p:
$P(X=n) = \binom{N}{n} p^n (1-p)^{N-n},\ n = 0, \ldots, N,$ where $\binom{N}{n} = \frac{N!}{(N-n)!n!}$
$E[X] = Np,\ \text{var}(X) = Np(1-p)$

3) Poission with parameter λ:
$P(X=n) = \frac{\lambda^n}{n!} e^{-\lambda},\ n \geq 0$
$E[X] = \lambda,\ \text{var}(X) = \lambda$

2. Continuous Random Variables

1) Uniformly distributed in $[a,b]$, for some $a < b$:
$f_X(x) = \frac{1}{b-a}$ where $a \leq x \leq b$
$E[X] = \frac{a+b}{2},\ \text{var}(X) = \frac{(b-a)^2}{12}$

2) Exponentially distributed with rate $\lambda > 0$:
$f_X(x) = \lambda e^{-\lambda x}$ where $x \geq 0$
$E[X] = \lambda^{-1},\ \text{var}(X) = \lambda^{-2}$

3) Gaussian, or normal, with mean μ and variance σ^2:
$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}$
$E[X] = \mu,\ \text{var} = \sigma^2$

Problem 1. (a) (2 points each, 8 points total. **You must provide brief explanations to justify your answers to get credit on all parts.**)

(i) Recall that the median, M, of the distribution of a random variable X is such that $P(X \leq M) = \frac{1}{2}$. Find the median of an exponential random variable X with rate λ.

Solution: We have $P(X \leq M) = \int_0^M \lambda e^{-\lambda x} = -e^{-\lambda M} + 1$. Thus, $e^{-\lambda M} + 1 = \frac{1}{2}$. Solving gives $M = \frac{\ln 2}{\lambda}$

(ii) **True/False** For events A, B, C, if $P(A|C)P(B|C) = P(A, B|C)$, then A and B are independent.

Solution: False. Consider the situation where you have a fair coin and a biased coin. Let C be the event that you pick the fair coin, A be the result of the first toss and B be the result of the second toss. Clearly, A and B are conditionally independent given C. However, A and B are not unconditionally indpendent.

(iii) What is a prefix code?

Solution: A code in which no codeword is a prefix to any other codeword. For example the code $\{1, 11\}$ is not a prefix code, while $\{0, 11\}$ is.

(iv) **True/False** Recall that in Lab 3, we used an ℓ-bit uniform-quantizer where we can only use $L = 2^\ell$ quantized values. If we model the error between a quantized signal and the original as a uniform random variable between 0 and $\frac{1}{2^\ell - 1}$, then the mean-squared-error will decrease linearly in ℓ.

Solution: False. It will decrease *exponentially*. Please see the solution of Lab 3 for details.

(b) (6 points) Recall that in a Binary Symmetric Channel (BSC), the input bit is flipped with probability p and received without error with probability $1-p$. Consider now cascading n BSCs such that the output of the first channel is fed to input of the second and so on, as shown in Figure 1.

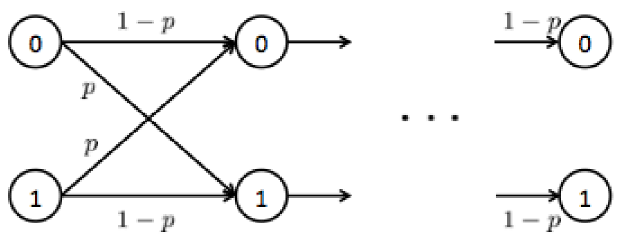

Figure 1: A cascaded BSC.

(i) (3 points) Find the probability that there were an even number of flips.
Solution: We have:

$$P(\text{even}) = \sum_{i=0}^{\lfloor \frac{n-1}{2} \rfloor} \binom{n}{2i} p^{2i}(1-p)^{n-2i}$$

$$P(\text{odd}) = \sum_{i=0}^{\lfloor \frac{n-1}{2} \rfloor} \binom{n}{2i+1} p^{2i+1}(1-p)^{n-2i-1}$$

Now, recall the binomial theorem: $(a+b)^n = \sum_{i=1}^{n} \binom{n}{i} a^i b^{n-i}$ and note that:

$$1 = (p + (1-p))^n = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = P(\text{even}) + P(\text{odd}) \tag{1}$$

Also, by the binomial theorem:

$$(-p + (1-p))^n = \sum_{i=0}^{n} \binom{n}{i} (-p)^i (1-p)^{n-i} = P(\text{even}) - P(\text{odd}) \tag{2}$$

Adding equations (1) and (2) gives: $2P(\text{even}) = 1 + (1-2p)^n$. And thus, $P(\text{even}) = \frac{1}{2}(1 + (1-2p)^n)$.

(ii) (3 points) Given that a 0 is received, what is the probability that a 0 was sent? Assume that a priori, the probability of sending a 0 is α, where $0 \le \alpha \le 1$.
Solution: Let S denote the symbol sent and R denote the symbol received. We have:

$$\begin{aligned} P(S=0|R=0) &= \frac{P(R=0|S=0)P(S=0)}{P(R=0)} \\ &= \frac{P(R=0|S=0)P(S=0)}{P(R=0|S=0)P(S=0) + P(R=0|S=1)P(S=1)} \end{aligned}$$

We note now that given $S = 0$, we will receive a 0 only if an even number of flips occurred, and we will receive a 1 if an odd number of flips occurred. Thus, we have:

$$\frac{P(R=0|S=0)P(S=0)}{P(R=0|S=0)P(S=0)+P(R=0|S=1)P(S=1)} = \frac{\alpha P(\text{even})}{\alpha P(\text{even}) + (1-\alpha)P(\text{odd})}$$

where $P(\text{even}) = \frac{1}{2}(1+(1-2p)^n)$ and $P(\text{odd}) = 1 - P(\text{even})$.

(c) (6 points) Let X and Y be independent random variables that are uniformly distributed on $[0, 1]$. Find $E[X|X < Y]$.

Solution: Clearly, given $X < Y$, the coordinate (X, Y) is uniformly distributed in the upper left half of the unit square (See Figure 2)

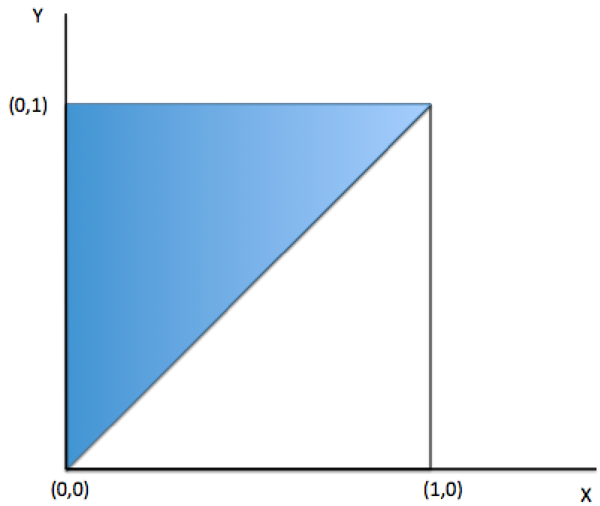

Figure 2: (X, Y) uniformly distributed in the shaded area given $X < Y$

Thus, we can see that conditioned on $X < Y$:

$$\begin{aligned} f_{X|X<Y}(x) &= \int_x^1 2dx \\ &= 2(1-x) \end{aligned}$$

so:

$$\begin{aligned} E[X|X<Y] &= \int_0^1 x f_{X|X<Y}(x)dx \\ &= \int_0^1 2x(1-x)dx \\ &= \frac{1}{3} \end{aligned}$$

(d) (6 points) Consider IID random variables $X_1, X_2, \ldots, X_5$ where $X_i \sim U(-1,1)$. Find $E[X_1 + X_2 + X_3 | X_1 + X_2 + X_3 + X_4 + X_5 = 2]$.

Solution: First notice that since X_i are IID random variables, $E[X_i | X_1 + X_2 + X_3 + X_4 + X_5 = 2] = E[X_j | X_1 + X_2 + X_3 + X_4 + X_5 = 2]$ where $i, j \in \{1,2,3,4,5\}$. Thus, noticing that $E[X_1+X_2+X_3+X_4+X_5 | X_1+X_2+X_3+X_4+X_5 = 2] = 2$, we can see that $E[X_1+X_2+X_3+X_4+X_5 | X_1+X_2+X_3+X_4+X_5 = 2] = 5(E[X_1 | X_1+X_2+X_3+X_4+X_5 = 2])$, so we have: $E[X_1 | X_1 + X_2 + X_3 + X_4 + X_5 = 2] = \frac{2}{5}$. Thus:

$$E[X_1+X_2+X_3 | X_1+X_2+X_3+X_4+X_5 = 2] = 3E[X_1 | X_1+X_2+X_3+X_4+X_5 = 2] = \frac{6}{5}$$

(e) (6 points) Consider two independent random variables X and Y that are both uniformly distributed on $[0,1]$. Let $U = \min(X,Y)$ and $V = \max(X,Y)$. Find $cov(U,V)$.

Solution: We are looking for $cov(U,V) = E[UV] - E[U]E[V]$. First notice that $UV = XY$, so $E[UV] = E[XY] = E[X]E[Y] = \frac{1}{4}$, where the second equality follows from the independence of X and Y. Now, to find $E[U]$ and $E[V]$, notice that randomly throwing two points on a line of length 1 is equivalent to throwing 3 points on a circle of circumference of radius 1, and letting one of the points be 0 (See Figure 3). On the circle, we can see that the segments between the points are identically distributed. Thus, since the three points break the circle into three segments, each of these segments will be equal in expectation, so they will all be of length $\frac{1}{3}$. Thus, $E[U] = \frac{1}{3}$ and $E[V] = \frac{2}{3}$. Putting this together gives $cov(U,V) = \frac{1}{36}$.

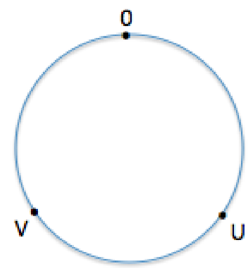

Figure 3: Unit line bent into a circle with $U, V, 0$ labeled.

Problem 2. (20 points) Consider the case of n graduate students who ride their bikes to their lab. Over the course of the day, they all forget which bike is theirs. When leaving, each graduate student takes a bike at random.

(a) (5 points) Let X be the number of graduate students that leave with their own bike. What is $E[X]$?

Solution: Let X_i be the event that student i leaves with his or her own bike. Note that $P(X_i = 1) = \frac{1}{n}$ for all the students, so we have: $E[X] = E[X_1 + X_2 + \cdots + X_n] = \sum_{i=1}^{n} E[X_i] = n\frac{1}{n} = 1$.

(b) (8 points) The situation is the same as above. Find $\mathrm{Var}(X)$.

Solution: Note that $\mathrm{Var}(X) = E[X^2] - E[X]^2$. We know from part a. that $E[X] = 1$, so we need to find $E[X^2]$.

$$\begin{aligned} E[X^2] &= E[(\sum_{i=1}^{n} X)^2] \\ &= E[\sum_{i=1}^{n} X_i^2 + \sum_{i \neq j} X_i X_j] \\ &= \sum_{i=1}^{n} E[X_i^2] + \sum_{i \neq j} E[X_i X_j] \end{aligned}$$

Now, notice that $P(X_i X_j = 1) = P(X_i, X_j = 1) = P(X_i = 1 | X_j = 1)P(X_j = 1) = \frac{1}{n} \cdot \frac{1}{n-1} = \frac{1}{n(n-1)}$. Also, we see that $E[X_i^2] = \frac{1}{n}$, so:

$$\begin{aligned} E[X^2] &= \sum_{i=1}^{n} \frac{1}{n} + \sum_{i \neq j} \frac{1}{n(n-1)} \\ &= 2 \end{aligned}$$

So $\mathrm{Var}(X) = 1$.

Now suppose that the bikes have unique locks on them and each student has a key to his or her own bike, but each student has forgotten which bike is theirs. Some graduate students have decided on the following solution: All the graduate students leave at the same time at the end of the day. Simultaneously, each student picks a bike uniformly at random, tries to unlock it, and leaves if successful. The remaining students pool the remaining bikes and begin another round. In each round, the remaining students pick one of the remaining bikes uniformly at random, leaving if they are able to unlock the bike. This continues until all students have left with their correct bikes. Let R_n be the random variable representing the number of rounds necessary for all n students to leave with the correct bike.

(c) (4 points) Find a recursive equation for $E[R_n]$ involving $E[R_1], E[R_2], \ldots, E[R_n]$.

Solution: Let N_1 be the number of students that leave with their own bike in the first round. We have by the tower property that $E[R_n] = E[E[R_n|N_1]] = \sum_{i=0}^{n} E[R_n|N_1 = i]P(N_1 = i)$. Here, we note that $E[R_n|N_1 = i] = 1 + E[R_{n-i}]$, where R_{n-i} is the number of rounds neccessary for $n-i$ graduate students to leave with the correct bike. We have:

$$\begin{aligned}
E[R_n] &= \sum_{i=0}^{n}(1 + E[R_{n-i}])P(X_1 = i) \\
&= 1 + \sum_{i=0}^{n} P(X_1 = i)E[R_{n-i}] \\
&= 1 + P(X_1 = 0)E[R_n] + \sum_{i=1}^{n} P(X_1 = i)E[R_{n-i}]
\end{aligned}$$

(d) (3 points) Find $E[R_n]$.

Solution: Intuitively, since the answer to part a. is 1, we claim that $E[R_n] = n$, and we show this by strong induction. For the base case R_1, we note that $E[R_1] = 1$ and we are done. Now we assume for all k such that $1 \le k \le n-1$, that $E[R_k] = k$. From part c we have:

$$\begin{aligned}
E[R_n] &= 1 + P(X_1 = 0)E[R_n] + \sum_{i=1}^{n} P(X_1 = i)(n - i) \\
&= 1 + P(X_1 = 0)E[R_n] + n(1 - P(X_1 = 0)) - \sum_{i=1}^{n} iP(X_1 = i) \\
&= 1 + P(X_1 = 0)E[R_n] + n(1 - P(X_1 = 0)) - E[X_1] \\
&= P(X_1 = 0)E[R_n] + n(1 - P(X_1 = 0))
\end{aligned}$$

We thus see that $E[R_n] = P(X_1 = 0)E[R_n] + n(1 - P(X_1 = 0))$. Rearranging gives: $E[R_n] = n$ and we are done.

Problem 3. (15 points) Two points are picked uniformly at random in the interval $[0, L]$.

(a) (8 points) Let the points be X_1, X_2 such that $0 \leq X_1 \leq X_2 \leq L$ as shown in Figure 2. Find the CDF of $X_2 - X_1$.

Figure 4: X_1 and X_2.

Solution: Let the original points (before relabeling) be A and B. Note that the CDF of $X_2 - X_1$ is exactly the same as the CDF of $|A - B|$. Thus, we may draw the area of interest (see the blue shaded region in Figure 5)

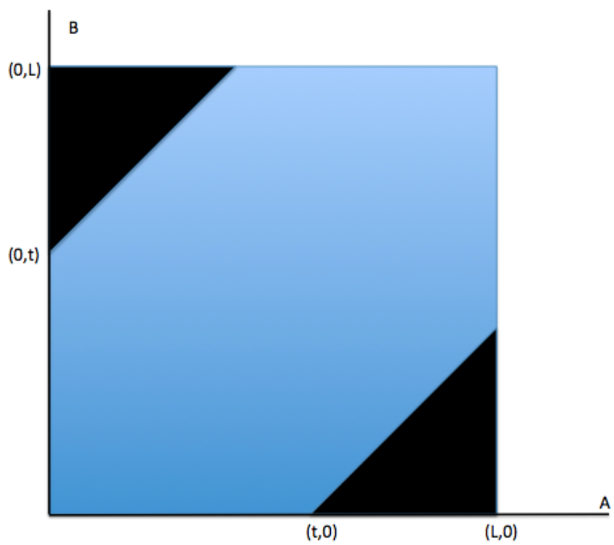

Figure 5: X_1 and X_2.

We see that

$$P(X_2 - X_2 \leq t) = P(|A - B| \leq t) = \frac{(L^2 - (L-t)^2)}{L^2} = 1 - \frac{(L-t)^2}{L^2}$$

(b) (7 points) What is the probability that a triangle can be formed from the lengths X_1, $X_2 - X_1$, and $L - X_2$?

Solution: We should have $X_1 + X_2 - X_1 > L - X_2$, $X_1 + L - X_2 > X_2 - X_1$, and $L - X_2 + X_2 - X_1 > X_1$. After simplifying, we get that the feasible region to have a triangle is $X_2 > L/2$, $X_1 < L/2$ and $X_2 - X_1 < L/2$. The pair (X_1, X_2) are uniformly distributed in the triangle $0 \leq X_1 \leq X_2 \leq L$ with area $L^2/2$. The feasible region has area $L^2/8$. Thus, the probability of having a triangle is $1/4$.

Problem 4. (12 points) Consider a 3-alphabet source X, with distribution as shown below, whose entropy $H(X) = 0.802$ bits per symbol.

P(X)	X
0.1	A
0.2	B
0.7	C

(a) (4 points) Find the average number of bits per symbol for encoding X using a Huffman code.

Solution: We may draw the Huffman tree and see that a A will have codeword of length 2, B will have a codeword of length 2, and C will have a codeword of length 1. Thus, the average number of bits per symbol is $2(0.1 + 0.2) + (0.7) = 1.3$.

(b) (6 points) Suppose now that a Huffman code is constructed for an alphabet consisting of blocks of symbols of X, with block size 2. In other words, each symbol is now a concatenation of two symbols from X. Find the average number of bits per symbol for this Huffman code.

Solution: Again, we may draw the Huffman tree, with the concatenated codewords and probabilities as follows:

P(X)	X
0.02	AB
0.02	BA
0.07	AC
0.07	CA
0.01	AA
0.14	BC
0.14	CB
0.04	BB
0.49	CC

Thus, a Huffman code is the following:

Length	Blocks
6	AA, AB
5	BA
4	BB, AC, CA
3	BC, CB
1	CC

(Note that this Huffman code is *not* unique). We can find the average bits per symbol to be $\frac{1}{2}[6(0.01 + 0.02) + 5(0.02) + 4(0.04 + 0.07 + 0.07) + 3(0.14 + 0.14) + 0.49] = 1.165$

(c) (2 points) Let the alphabet consist of blocks with block size going to ∞. What is the average number of bits per symbol?

Solution: Recall from lecture and lab 3 that as block size goes to ∞, the average bits per symbol will go to $H(X) = 0.802$.

Problem 5. (25 points, 5 points for each part) In this problem, we will consider a matrix $A \in \mathbb{R}^{2n \times n}$, where the entries of A are IID random variables. The entry of A in the position (i, j) is denoted by $A_{i,j}$, and it is distributed as follows:

$$A_{i,j} = \begin{cases} 0 & \text{w.p. } 0.5 \\ 1 & \text{w.p. } 0.25 \\ -1 & \text{w.p. } 0.25 \end{cases}.$$

We further define U (for upper) to be a matrix consisting of the first n rows of the matrix A, and L (for lower) to be a matrix consisting of the last n rows of the matrix A. Thus, it is clear that $U, L \in \mathbb{R}^{n \times n}$. See Fig. 6 below.

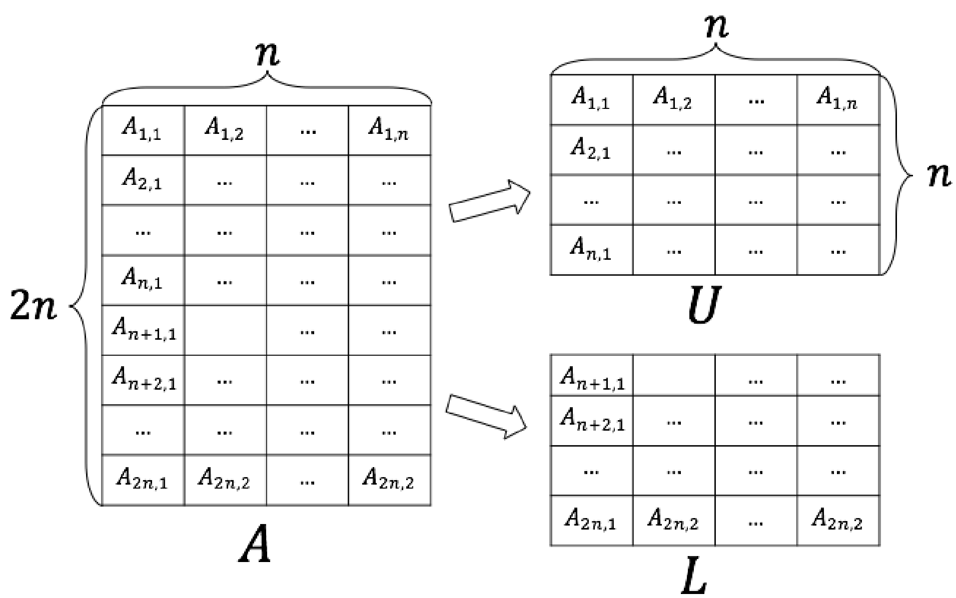

Figure 6: A, U, and L

The **density** of a matrix is defined as the number of non-zero elements in the matrix.

(a) What is the expected density of the matrix U?

Solution: Each element is non-zero with probability 0.5, so it is $\frac{n^2}{2}$.

(b) What is the expected density of the matrix sum of U and L, i.e., of the matrix $W = U + L$ (where '$+$' denotes real-valued addition)?

Solution: Each element of $U + L$ zero if the two corresponding elements from U and L are both zero or negative of each other. This happens with probability $\frac{1}{4} + 2\frac{1}{16} = \frac{3}{8}$. Thus the probability of a nonzero element is $\frac{5}{8}$, so the expected density is $\frac{5n^2}{8}$.

Alice needs to compute a matrix multiplication Ax, $x \in \mathbb{R}^n$, for her homework assignment. Since the matrix A is too large, she wants to compute Ax in *parallel* using two machines. Note that $Ax = \begin{bmatrix} U \\ L \end{bmatrix} x = \begin{bmatrix} Ux \\ Lx \end{bmatrix}$. Thus, one can simply compute Ux using one machine and compute Lx using the other. Once these two computations are done, one can simply concatenate them to obtain Ax.

(c) Assume the performance of both machines is unpredictable. Denote the time to compute Ux as T_U, and that to compute Lx as T_L. The T_U and T_L are exponentially distributed with rate 1. Thus, the waiting time to obtain Ax is $\max(T_U, T_L)$. See Fig. 7 for illustration.

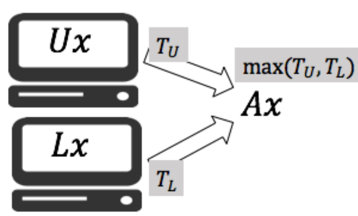

Figure 7: Parallel computing scheme for part c.

Find the expected time to obtain Ax, i.e., $E[\max(T_U, T_L)]$.

Solution: Consider $M = \max(T_U, T_L)$. We have $P(M \leq t) = P(T_u \leq t, T_L \leq t) = (1 - e^{-t})^2$. Thus, we can see that $f_M(m) = 2e^{-t} - 2e^{-2t}$. We note that finding the expected value of M will then come down to twice the expected value of an exponential random variable with rate 1 minus the expected value of an exponential random variable with rate 2. Thus, the expected time to obtain Ax is $\frac{3}{2}$.

Alice realizes that the expected waiting time to compute Ax is so large that she will not be able to submit her solution on time. She borrows two extra machines from her friend Bob, but how should she use her 4 machines to be maximally efficient?

(d) Donald, a fellow student, suggests the idea of replicating each tasks on 2 machines as shown below in Fig. 8.

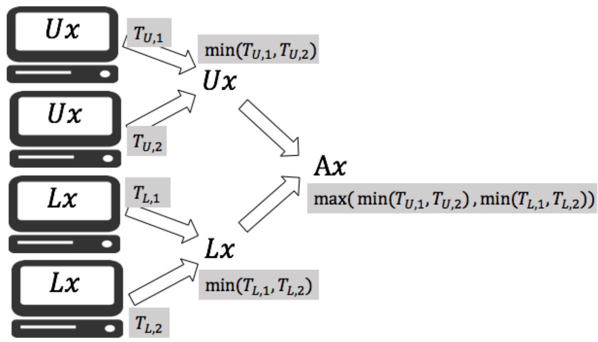

Figure 8: Parallel computing scheme for part d.

Hence, the waiting time is now $\max(\min(T_{U,1}, T_{U,2}), \min(T_{L,1}, T_{L,2}))$. Find your expected waiting time, i.e., $E[\max(\min(T_{U,1}, T_{U,2}), \min(T_{L,1}, T_{L,2}))]$.

Solution: Let $T_U = \min(T_{U,1}, T_{U,2})$ and $T_L = \min(T_{L,1}, T_{L,2})$. Note that T_U, T_L are both exponentially distributed with rate 2, and are independent. Now, let $M = \max(T_U, T_L)$ and note that now $f_M(t) = 4e^{-2t}(1-e^{-2t})$. Thus, we see that $E[M]$ is equivalent to twice the expectation of an exponential random variable with rate 2 minus the expectation of an exponential random variable with rate 4, and $E[\max(\min(T_{U,1}, T_{U,2}), \min(T_{L,1}, T_{L,2}))] = \frac{3}{4}$.

(e) Bernie, another fellow student, suggests the following alternative: the first machine computes Ux, and the second machine computes Lx; the third machine computes $(U+L)x$, and the fourth machine computes $(U-L)x$. As you can easily deduce, with this scheme, one can compute Ax based on *any* 2 out of the 4 machines producing their results. *(In case you're confused, don't worry. You are not required to understand why this is true.)* That is, the waiting time under this scheme is the 2nd smallest value of T_U, T_L, T_{U+L}, and T_{U-L}. See Fig. 9 for illustration.

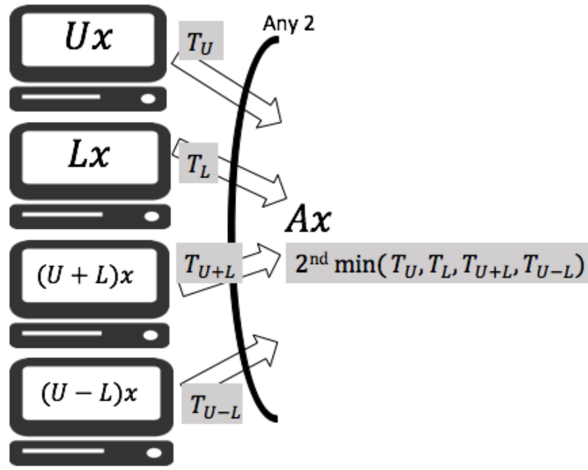

Figure 9: Parallel computing scheme for part e.

However, since $U+L$ and $U-L$ are "denser" (i.e. have fewer zero entries) than U or L, their computing time takes longer time on average. More precisely, assume that the time to compute $(U+L)x$, denoted by T_{U+L}, is randomly distributed as an exponential random variable with rate 0.75, and the same for T_{U-L}. Find your expected waiting time. Is Bernie's scheme better than Donald's scheme?

Solution: The first waiting time is the minimum of 4 exponential timers, and the total rate is $2{+}1.5 = 3.5$. Thus, $1/3.5$ is the first time to see any output to be done. Then, with probability $2/3.5$, you have to wait for another $1/2.5$, and with probability $1.5/3.5$, you have wait for another $1/2.75$. Thus, $1/3.5{+}2/(3.5{*}2.5){+}1.5/(3.5{*}2.75) = 258/385 \simeq 0.67$, which is 10.7% faster than Bernie's scheme.

END OF THE EXAM.

Please check whether you have written your name and SID on every page.