CS Probability and Statistics for Computer Scientists
Spring, 2025

Name (Print last name first): ________________________________

For decimal fractions, show at least four digits, rounding correctly.

Do not simply memorize the question and solution. Understand the questions and understand the principles in the solutions.

Question 1: 10 Points

Prove

$$A^C = (A^C \cap B) \cup (A^C \cap B^C)$$

$$\begin{aligned} A^C &= A^C \cap \Omega \\ &= A^C \cup (B \cap B^C) \\ &= (A^C \cup B^C) \cap (A^C \cap B^C) \end{aligned}$$

Question 2: Prove if $A \subset B$, $E \subset B$, $C \subset D$, and $F \subset D$ then prove the following statement:

$$(A \cap E) \times (C \cap F) \subset B \times D$$

- Let $(a, c) \in (A \cap E) \times (C \cap F)$
- Then $a \in A \cap E$ so $a \in A$
- Then $c \in C \cap F$ and $c \in C$
- Then $a \in B$ and $c \in D$
- Then $(a, c) \in B \times D$

Question 3: 10 points

A six sided die is loaded in a way that the two is twice as likely as the one, three is three times as likely, four is four times as likely, five is five times as likely, and six is six times as likely.

(a) Construct the PMF for the probability function for the values on the die.

(b) What is the probability that the outcome is greater than four?

(a)

1	$\frac{1}{21}$
2	$\frac{2}{21}$
3	$\frac{3}{21}$
4	$\frac{4}{21}$
5	$\frac{5}{21}$
6	$\frac{6}{21}$

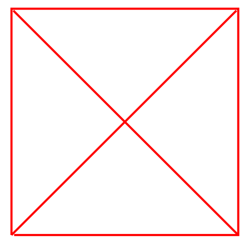
(b) $\frac{11}{21}$

Question 4:

Let X, Y be uniform random variables, both in [0,1].

What is $P(X + Y \leq 1 \text{ and } Y - X < 0)$?

Answer: ....

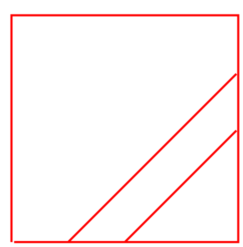


$P(X + Y \leq 1 \text{ and } Y - X < 0) = \frac{1}{4}$
Partial credit: 1 point for drawing the picture.

Question 5: Let X, Y be uniform random variables, both in [0,1], let $Z = X - Y$

What is $P\left(\frac{1}{4} \leq Z \leq \frac{1}{2}\right)$?

Answer: ....



$P\left(\frac{1}{4} \leq Z \leq \frac{1}{2}\right) = \frac{1}{2} \times \frac{3}{4} \times \frac{3}{4} - \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{9}{32} - \frac{1}{8} = \frac{5}{32}$
Partial credit: 1 point for drawing the picture.

Question 6:

Suppose that X is a random variable with integer values uniform distribution in the interval [0,10].

2 points: What is $E[X]$?
Answer: 5

You wonder what $E[min(X_1, X_2)]$ is, where X_1 and X_2 are two events from X. Write a pseudo-code program to determine $E[min(X_1, X_2)]$

```
import math\\
import random\\

minsum = 0\\

n = 100000\\

for c1 in range(n):\\
        a = random.random()\\
        b = random.random()\\

        minsum += min(a,b)\\

print('E[min(a,b)]= ' + str(minsum/n))\\
```

You could also have a deterministic solution where you check every possible pair.

Question 7: In the town Imaginaryville yearly precipitation is modeled as a normal random variable with mean 360 cm and standard deviation 30 cm.

Answers involving Φ *may be left as function calls without using any function lookup. So,* $\Phi(.5)$ *could be a correct answer. Note that the argument in* $\Phi()$ *should be non-negative.*

(a) (3 points) What is the probability in a randomly chosen year that the precipitation R is below 320 cm?

$$P(X < x) = \Phi\left(\frac{x - \mu_X}{\sigma_X}\right)$$

$$P(R < 320) = P\left(\frac{R-360}{30} < \frac{320-360}{30}\right) = P\left(\frac{R-360}{30} < -\frac{4}{3}\right)$$
$$= \Phi\left(-\frac{4}{3}\right) = 1 - \Phi\left(\frac{4}{3}\right)$$

(b) (3 points) What is the probability in a randomly chosen year that the precipitation is above 375 cm?

$$P(R > 375) = 1 - P(R < 375) = 1 - P\left(\frac{R-360}{30} < \frac{375-360}{30}\right)$$
$$= 1 - P\left(\frac{R-375}{30} < \frac{1}{2}\right) = 1 - \Phi\left(\frac{1}{2}\right)$$

(c) (4 points) What is the probability in a randomly chosen year that the precipitation is between 320 cm and 350 cm?

$$P(275 < R < 350) = P(R < 350) - P(R < 320)$$
$$= P\left(\frac{R-360}{30} < \frac{350-360}{30}\right) - P\left(\frac{R-360}{30} < \frac{320-360}{30}\right)$$
$$= \Phi\left(\frac{350-360}{30}\right) - \Phi\left(\frac{320-360}{30}\right)$$
$$= \Phi\left(-\frac{1}{3}\right) - \Phi\left(-\frac{4}{3}\right) = \Phi\left(\frac{4}{3}\right) - \Phi\left(\frac{1}{3}\right)$$

Question 8: 10 Points

Metropolis has three redundant roads connecting with Mountain City. Two roads connect Mountain City with Small Town.

Road one from Metropolis to Mountain City is open with probability .8, road two from Metropolis to Mountain City is open with probability .6, and road three from Metropolis to Mountain City is open with probability .7. Also, road one from Mountain City to Small Town is open with probability .8 and road two from Mountain City to Small Town is open with probability .9.

What is the probability that there will be at least one set of roads from Metropolis to Small Town?

HINT: Draw a diagram.

Answer: ....

P(Metropolis to Small Town open) =
P(Metropolis to Mountain City open) × P(Mountain City to Small Town open) =
(1 - P(Metropolis to Mountain City closed))× (1 - P(Mountain City to Small Town closed)) =
$(1-(1-.8)\times(1-.6)\times(1-.7))\times(1-(1-.8)\times(1-.9))$
$= (1-.2\times.4\times.3)\times(1-.2\times.1)$
$= (1-.024)\times(1-.02)$
$.976\times.98$
=**.95648** or **.9565**

Formula correct **6**
Arithmetic correct **2**
Answer without work **1**

Question 9:

You observe that the time to commute to work depends on when you leave your apartment. Over two weeks you record the time you leave and the time you arrive at work.

Leave	Arrive
7:05	7:36
7:10	7:30
7:10	7:35
7:15	7:27
7:15	7:31
7:15	7:29
7:20	7:35
7:20	7:27
7:25	7:33
7:25	7:31
7:30	7:37

(a) [3 points] Does it appear that there is a relation between time you left and travel time to get to work?

Yes

(b) [5 points]Would you estimate the correlation ρ between the two data sets would be (circle one):

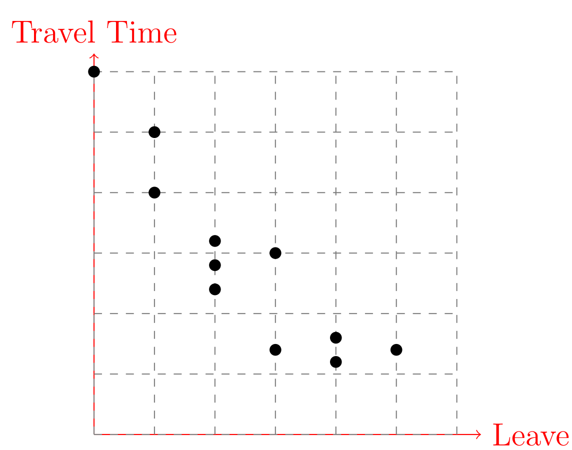
i. $-1 \leq \rho \leq -.5$

ii. $-.25 \leq \rho \leq .25$

iii. $.5 \leq \rho \leq 1$

$\rho \leq -.25$

Actual correlation is $-.899$.



Question 10:

All answers should be left in terms of Φ.

Let X and Y be independent Standard Normal Distribution random variables

(a) [2 points] What is the probability that $|X| < .75$?

(b) [2 points] What is the probability that $-.1 < Y < 0.6$?

(c) [3 points] What is μ_{2X-5Y}?

(d) [3 points] What is σ_{5X+12Y}?

a) $P(-.75 < X < .75) = \Phi(.75) - \Phi(-.75) = 2\Phi(.75) - 1$
b) $\Phi(0.6) - \Phi(-.1) = \Phi(0.6) + \Phi(.1) - 1$
c) 0
d) 13

Question 11:

Let X and Y be independent Standard Normal Distribution random variables

(a) [2 points] What is the probability that $|X| < .75$?

(b) [2 points] What is the probability that $-.1 < Y < 0.6$?

(c) [3 points] What is the probability that $0 < 3X + 4Y < \sqrt{20}$?

(d) [3 points] What is the probability that $0 < 3X - 4Y < \sqrt{20}$?

a) $P(-.75 < X < .75) = \Phi(.75) - \Phi(-.75) = 2\Phi(.75) - 1$
b) $\Phi(0.6) - \Phi(-.1) = \Phi(0.6) + \Phi(.1) - 1$
c) $\Phi\left(\frac{\sqrt{20}}{5}\right) - \frac{1}{2}$
d) $\Phi\left(\frac{\sqrt{20}}{5}\right) - \frac{1}{2}$

Question 12: 10 Points

Ming rides to school on his bike or takes the bus, depending on the weather. On sunny days he rides the bike, and the trip takes 20-30 minutes, with uniform probability, depending on traffic lights and other delats. On rainy days he rides the bus, taking 30-50 minutes.

If the probability of a rainy day is .25, what is Ming's expected trip time?

X is between 20 and 30 minutes **uniformly** with probability $\frac{3}{4}$, and between 30 and 50 minutes **uniformly** with probability $\frac{1}{4}$.

$$f_X(x) = \begin{cases} c_1 & \text{if } 20 \le x \le 30\} \\ c_2 & \text{if } 30 \le x \le 50\} \\ 0 & \text{otherwise} \end{cases}$$

$$P(sunny) = \frac{3}{4} = \int_{20}^{30} f_X(x)\,dx = \int_{20}^{30} c_1\,dx = 10c_1$$

$$c_1 = \frac{3}{40}$$

$$P(rainy) = \frac{1}{4} = \int_{30}^{50} f_X(x)\,dx = \int_{30}^{50} c_2\,dx = 20c_2$$

$$c_2 = \frac{1}{80}$$

$$f_X(x) = \begin{cases} \frac{3}{40} & \text{if } 20 \le x \le 30\} \\ \frac{1}{80} & \text{if } 30 \le x \le 50\} \\ 0 & \text{otherwise} \end{cases}$$

$$E[X] = \int_{20}^{50} x f_X(x)\,dx = \int_{20}^{30} x\frac{3}{40}\,dx + \int_{30}^{50} x\frac{1}{80}\,dx$$

$$= \frac{3}{80}\times(900-400)+\frac{1}{160}\times(2500-900) = \frac{1500}{80}+\frac{1600}{160} = 18.75+10 = 28.75$$

This can actually be done more easily,

$$E[bike]\times\frac{3}{4} + E[bus]\times\frac{1}{4} = 25\times\frac{3}{4} + 40\times\frac{1}{4} = 28.75$$

Question 13: 10 Points

Let X be a random variable that takes integer values from 0 to 9 with equal probability $\frac{1}{10}$.

(a) Find the PMF of the random variable $Y = X \mod (3)$.
Hint: The mod function gives the remainder of X divided by 3

X	$Y = X \mod (3)$
0	0
1	1
2	2
3	0
4	1
5	2
6	0
7	1
8	2
9	0

Y	Count	P(Y)
0	4	.4
1	3	.3
2	3	.3

(b) Find the PMF of the random variable $Z = 5 \mod (X + 1)$

X	$X + 1$	$Z = 5 \mod (X + 1)$
0	1	0
1	2	1
2	3	2
3	4	1
4	5	0
5	6	5
6	7	5
7	8	5
8	9	5
9	10	5

Y	Count	P(Y)
0	2	.2
1	2	.2
2	1	.1
5	5	.5
else		0

Question 14: 10 Points

Let X be a discrete random variable with PMF

$$p_X(x) = \begin{cases} \frac{x^2}{a} & \text{if } x \in \{-3,-2,-1,0,1,2,3\} \\ 0 & \text{otherwise} \end{cases}$$

Find a and $E[X]$

a: ....

$E[X]$: ....

Since

$$1 = \sum_x P_X(x) = \frac{1}{a} \sum_{x=-3}^{3} x^2$$

$$a = \sum_{x=-3}^{3} x^2 = 28$$

Since X is symmetric around 0, $E[X] = 0$

Question 15: 10 Points

Two random variables X and Y have the joint distribution function, $P_{X,Y}(0,0) = 0.2, P_{X,Y}(0,2) = 0.1, P_{X,Y}(0,3) = .1, P_{X,Y}(1,1) = 0.1,$ $P_{X,Y}(1,3) = .1, P_{X,Y}(2,0) = 0.1, P_{X,Y}(2,2) = 0.1, P_{X,Y}(3,1) = .1,$ $P_{X,Y}(3,2) = .1$, and $P_{X,Y}(x,y) = 0$ for all other pairs (x,y).

(a) Find the probability mass function of $Z = X + Y$.

(b) Find the probability mass function of $U = X - Y$.

(c) Find the probability mass function of $V = XY$.

(d) Are X and Y independent? Why or why not?
Yes No

No. $P(X = 0 \; and \; Y = 0) = .2 \neq .4 \times .3 = P(X = 0) \times P(Y = 0)$

Hint: draw a table of values first.

Y

		0	1	2	3
	0	.2		.1	.1
X	1		.1		.1
	2	.1		.1	
	3		.1	.1	

#	Z = X + Y
0	.2
1	0
2	.3
3	.1
4	.3
5	.1
6	0

#	U = X - Y
-3	.1
-2	.2
-1	.0
0	.4
1	.1
2	.2
3	.0

#	V = X * Y
0	.5
1	.1
2	0
3	.2
4	.1
6	.1
9	0

Omitting rows for 0 values is fine.

Question 16:

Let X be a random variable with PDF

$$f_X(x) = \begin{cases} 0 & x < 0 \\ cx^2 & 0 \le x \le 2 \\ 0 & x > 2 \end{cases}$$

(a) What is c?

(b) What is $E(X)$?

a)

$$\begin{aligned} 1 &= \int_0^2 cx^2 \, dx \\ &= c \int_0^2 x^2 \, dx \\ &= c\frac{1}{3}x^3|_0^2 \\ &= c\frac{8}{3} \\ c &= \frac{3}{8} \end{aligned}$$

b)

$$\begin{aligned} E(X) &= \int_0^2 \frac{3}{8} x x^2 \, dx \\ E(X) &= \int_0^2 \frac{3}{8} x^3 \, dx \\ &= \frac{3}{32} x^4|_0^2 = 1.5 \end{aligned}$$

Question 17:

The joint PDF of two continuous random variables X and Y is given by

$$\begin{cases} 4xy, & \text{if } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

(a) Find the marginal PDF of X and Y, i.e., $f_X(x)$ and $f_Y(y)$

(b) Are X and Y independent?

a) $f_X(x) = 2x$, if $0 \leq x \leq 1$, $f_Y(y) = 2y$, if $0 \leq y \leq 1$

b) Yes

Question 18:

Passwords should be chosen from a set of symbols containing 26 lower case alphabetic letters, 26 upper case alphabetic letters, 10 digits, and special characters $\$, \#, @, \%$.

How many passwords are possible if:

(a) 5 Points: the password must have exactly eight characters, at least one upper case letter, at least one lower case letter, at least one special character, and at least one digit without repeating characters.

Hint: Just write the formula to compute the number; don't compute the actual number.

Answer: ....

(b) 5 Points: the password must have exactly eight characters, at least one upper case letter, at least one lower case letter, at least one special character, and at least one digit with repeating characters allowed.

Hint: Just write the formula to compute the number; don't compute the actual number.

Answer: ....

$26 \times 26 \times 4 \times 10 \times \binom{62}{4} \times 8!$

$26 \times 26 \times 10 \times 4 \times \binom{8}{4} \times 4! \times 66^4$

-2 points for omitting 8!.

Question 19:

Suppose Apple notebook comes in either Black or White. Also, there are two tiers of configurations: the basic one is called MacBook, and high end one MacBook Pro. Assume that an apple store has a total of 100 white models and 20 black models. Out of the 100 white models, 30 of them are MacBook Pro while out of the 20 black models, 10 are MacBook Pro. A customer walks into a store and wants to check out a notebook. The saleperson randomly opened a box and it turns out to be a MacBook Pro. Find out the probability that this is a black model.

Let A be the event that MacBook Pro is chosen and B be the event such that the computer is a black model. Then $P(B|A)$ is the desired probability.

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)} = \frac{\frac{1}{2} \times \frac{20}{100+20}}{\frac{10+30}{100+20}} = \frac{1}{4}$$

Question 20:

You have been invited to play a game of darts with a peculiar dartboard. You will pay your opponent \$3.00 and throw a dart, then he will pay you the amount shown in the sector of the board where you hit. You are confident of your ability to hit the board every throw, but you think you will hit the board in a random spot.

The board is square, and is divided into four squares by a vertical line bisecting the square and a horizontal line bisecting the square. The upper left square contains "\$4". The lower right square contains "\$3". The lower left square contains "\$2". The upper right square contains "\$1".

Hint: Draw a picture.

(a) What is the expected value of a throw?

2.5

(b) If you throw numerous times, do you expect to

Lose money? Make money? Break even?

Lose money

Question 21:

A fire station is to be built along a street of length L. If fire will occur at points uniformly chosen on $(0, L)$, where should be the station be located so as to minimize the expected distance from the fire? That is, choose a so as to minimize $E[|X - a|]$, where X is uniformly distributed over $(0, L)$.

$$\begin{aligned} E[|X-a|] &= \int_0^a (a-u)f(u)du + \int_a^L (u-a)f(u)du \\ &= \int_0^a (a-u)\frac{1}{L}du + \int_a^L (u-a)\frac{1}{L}du \\ &= \frac{1}{L}\left(a^2 - \frac{a^2}{2} + \frac{L^2}{2} - aL - \frac{a^2}{2} + a^2\right) \\ &= \frac{L}{2} + \frac{a^2}{L} - a \end{aligned}$$

which is minimized at $a = \frac{L}{2}$.

Question 22:

Let an experiment be tossing two six faced dice. Let A be the event such that the outcome is a double (i.e., two identical numbers). Let B be the event that the sum of the two is 8, 9, or 10. Are A and B independent events?

Hint: count carefully.

Yes.
$A = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$,
$B = \{(2,6),(3,5),(3,6),(4,4),(4,5),(4,6),(5,3),(5,4),(5,5),(6,2),(6,3),(6,4)\}$
Thus, $P(A) = \frac{6}{36} = \frac{1}{6}$, $P(B) = \frac{12}{36} = \frac{1}{3}$, $P(A \cap B) = \frac{2}{36} = \frac{1}{18}$
Since $P(A)P(B) = P(A \cap B)$, they are independent.

Question 23:

X is a uniform random variable, $5 \leq X \leq 10$. Y is a normal random variable with mean 2 and standard deviation 1.

Write a pseudo code program to estimate the expected value of $X^2 + Y^3$

```
sum = 0
loop 10000
   X = 5*rand()+5
   Y = rand.normal (2,1)
   sum= sum + (X^2+Y^3)

print sum/10000
```

Question 24:

I want to paint the interior of the house using the same color in every room.

I calculated that the total area to cover is 2700 square feet, and the paint I want to use specifies that a gallon covers 400 square feet. The standard deviation is 40 square feet per gallon.

I want to compute the probability that I will have enough paint if I get 7 gallons.

What is the probability that 7 gallons will cover 2700 square feet.?

Hint: the formula in the Central Limit Theorem for normal distributions with mean μ and standard deviation σ is:

$$F_{Z_n}(z) = \mathbf{P}\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \frac{z - n\mu}{\sigma\sqrt{n}}\right\} \to \Phi\left(\frac{z - n\mu}{\sigma\sqrt{n}}\right)$$

$$\begin{aligned}
F_{Z_n}(z) &= 1 - \mathbf{P}\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \frac{z - n\mu}{\sigma\sqrt{n}}\right\} \\
&= 1 - \mathbf{P}\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \frac{2700 - 7 \times 400}{40\sqrt{7}}\right\} \\
&= 1 - \Phi\left(\frac{2700 - 7 \times 400}{40\sqrt{7}}\right) \\
&= 1 - \Phi\left(\frac{-100}{105.83}\right) \\
&= 1 - \Phi\left(-.9449\right) \\
&= \Phi\left(.9449\right)
\end{aligned}$$

Question 25:

Guitar strings should be manufactured with very little variation in thickness. Joe's Guitar String Company manufactures strings and has a testing process that detects problems with a reliability of 97%. That is, the test finds 97% of strings that vary more than .001 inch in thickness. It determines that the string varies less than .001 inch with a reliability of 98.9%. A much more expensive independent test finds that actual success rate in the manufacturing process is 99.8% of strings.

Suppose a particular string fails the test. That is, the test determines that the string is thicker or thinner than specification by more than .001.

What is the probability that the string is bad and should be rejected?

Hint: Use Baye's Rule for two events:

$$\mathbf{P}\{B|A\} = \frac{\mathbf{P}\{A|B\}\mathbf{P}\{B\}}{\mathbf{P}\{A|B\}\mathbf{P}\{B\} + \mathbf{P}\{A|\overline{B}\}\mathbf{P}\{\overline{B}\}}$$

Another hint: write down in words what the events A and B represent, as well as what is meant by $\mathbf{P}\{A\}$, $\mathbf{P}\{A|B\}$ and every other term in the formula. Then substitute those pieces in the formula.

$P(B)$ =.002 - string has a piece with diameter outside of range ±.002"

$P(\overline{B})$ =.998

$P(A)$ string failed test

$P(A|B)$ =.97 - string failed test, given that it actually out of spec

$P(\overline{A}|\overline{B})$ =.989 - string passes the test, if the string is good

$P(A|\overline{B})$ =.011 - string passes the test, if the string is good

$$\mathbf{P}\{B|A\} = \frac{.97 \times .002}{.97 \times .002 + .011 \times .998} = \mathbf{.1502}$$

Hints

- DeMorgan's Laws

$$(A \cap B)^C = A^C \cup B^C$$
$$(A \cup B)^C = A^C \cap B^C$$

- $E(X) = \sum_{x \in X} xp(x)$ or $E(X) = \int_{x \in X} xp(x)\, dx$

- Suppose A, B independent.

$$P(A\&B) = P(A \cap B) = P(A) \times P(B)$$

- Permutations of m things out of n things at a time

$$_nP_m = \frac{n!}{(n-m)!}$$

- Combinations of m things out of n things

$$_nP_m = \frac{n!}{m!(n-m)!}$$

-

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- For X, Y independent normal random variables

$$\mu_{X+Y} = \mu_X + \mu_Y$$
$$\sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2}$$
$$P(X < x) = \Phi\left(\frac{x - \mu_X}{\sigma_X}\right)$$

- $\Phi(x) = 1 - \Phi(-x)$

- $var(X) = E[X^2] - (E[X])^2$

- Central Limit Theorem for normal distributions with mean μ and standard deviation σ:

$$F_{Z_n}(z) = P\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \frac{z - n\mu}{\sigma\sqrt{n}}\right\} \to \Phi\left(\frac{z - n\mu}{\sigma\sqrt{n}}\right)$$

- $Cov(X, Y) = E[XY] - E[X]E[Y]$