Name: ____

ID: ____

On your desk you should have only the exam paper, writing tools, and the cheat-sheet. The cheat-sheet is one page handwritten on both sides.

The exams are color coded. Your exam should have different color than that of your neighbours to the left, right and in front.

There are 11 questions in this exam, totalling 125 points. The final score is determined by summing all the points and taking the min of the sum and 100. For a final grade of A+ it is necessary (but not sufficient) to get more than 100 points in the final.

Be clear and concise. Write your answers **in the space provided** after each question. Whenever possible, **write your answers as expressions, not as numbers**. If an expression is reused later in the problem, you are encouraged to assign it to a letter variable and use the variable in the later expressions. For example, suppose in the first part of the problem you find the answer to be $C(52,5)$ and that in a later part the answer is $22/C(52,5)$. You can define in the first part $a = C(52,5)$ and write the second expression as $22/a$. This will save you time and space and is easier for us to grade.

Do your work at the empty space. If your answer is incorrect, we will look at your work and give you credit if you were going in the right direction.

1	12pt.
2	9pt.
3	10pt.
4	10pt.
5	10pt.
6	9pt.
7	9pt.
8	9pt.
9	20pt.
10	11pt.
11	16pt.

Total 125

This exam was graded by:

abalsubr@eng.ucsd.edu iawwal@eng.ucsd.edu sjrao@eng.ucsd.edu zzhai@ucsd.edu yfreund@eng.ucsd.edu

1. (12 pts)

Let $X_1, X_2, \ldots, X_{100}$ be the outcomes of 100 independent tosses of a fair coin.
$P(X_i = 0) = P(X_i = 1) = 0.5$

1. (1 pt) $\mathbb{E}(X_1) =$ ____________________
2. (1 pt) $var(X_1) =$ ____________________

Define the random variable $X = X_1 - X_2$.

1. (1 pt) $\mathbb{E}(X) =$ ____________________
2. (1 pt) $var(X) =$ ____________________

Define the random variable $Y = X_1 - 2X_2 + X_3$.

1. (2 pts) $\mathbb{E}(Y) =$ ____________________
2. (2 pts) $var(Y) =$ ____________________

Define the random variable $Z = X_1 - X_2 + X_3 - X_4 + \ldots + X_{97} - X_{98}$.

1. (2 pts) $\mathbb{E}(Z) =$ ____________________
2. (2 pts) $var(Z) =$ ____________________

Correct Answers:

- `1/2`
- `1/4`
- `0`
- `1/2`
- `0`
- `6/4`
- `0`
- `98/4`

2. (9 pts)

For the following joint probability matrices, fill in the missing values and answer the questions.("Correlated" = either positively or negatively correlated.)

	X=1	X=2	X=3	Marginal over Y
Y=1	1/8	0	1/8	
Y=2		1/2	0	1/2
Y=3	1/8	0	1/8	1/4
Marginal over X	1/4	1/2		*

- (1pt)Are X and Y dependent? ________________
- (1pt)Are X and Y correlated? ________________
- (1pt)$P(X=2|Y=2) =$ ________________

	X=1	X=2	X=3	Marginal over Y
Y=1	1/10			1/5
Y=2	1/5	1/10	1/10	2/5
Y=3	1/5	1/10	1/10	
Marginal over X	1/2	1/4	1/4	*

- (1pt)Are X and Y dependent? ________________
- (1pt)Are X and Y correlated? ________________
- (1pt)$P(Y=2|X=3) =$ ________________

	X=1	X=2	X=3	Marginal over Y
Y=1	1/3	0	0	1/3
Y=2	0	1/6	0	1/6
Y=3		1/6	1/6	
Marginal over X	1/2		1/6	*

- (1pt)Are X and Y dependent? ________________
- (1pt)Are X and Y correlated? ________________
- (1pt)$P(X=3|Y=3) =$ ________________

Correct Answers:

- yes
- no
- 1
- no
- no
- 2/5
- yes
- yes
- 1/3

3. (10 pts)

You are given a card deck of 44 cards, consisting all combinations of 11 ranks and 4 suits. Suppose you draw a hand of 5 cards out of this deck, what is the probability that it will be a *straight* ?

We will refer to the ranks by numbers 1,2,3,4,... . The corresponding ranks in the standard deck are 1=Ace, 2=2,...,11=J, 12=Q,13=K

Recall that a *straight* refers to five cards whose ranks are in sequence, but whose suits are **not all the same** (otherwise it is called a straight flush). The 1 can be used either as a low card or a high card when forming a straight. For example, both (1,2,3,4,5) and (8,9,10,11,1) are valid straight ranks, but not (9,10,11,1,2).

1. (2 pts) The number of possibilities for the ranks of a straight is ________________ .
2. (3 pts) The suits can be anything other than all equal, so the number of possibilities for the suits of a straight is ________________ .
3. (3 pts) Thus the number of straight hands is ________________ .
4. (2 pts) Finally, the probability that a randomly chosen hand is a straight is ________________ .

Correct Answers:

- `11-3`
- `4**5 - 4`
- `(11-3)*(4**5-4)`
- `(11-3)*(4**5-4)/C(44,5)`

4. (10 pts)

You are given a card deck of 52 cards, consisting all combinations of 13 ranks and 4 suits. Suppose you have been dealt 4$\heartsuit$ and 5 $\heartsuit$. What is the probability that you will get a straight given that you have been dealt these two cards, and that the board is "3$\clubsuit$, Q$\spadesuit$, K$\diamondsuit$"?

With this flop we can make a straight with the set $\{6,7\}$, $\{2,6\}$ or $\{A,2\}$ on the remaining two cards to be dealt.

- (2 pts) The number of card pairs such that one is a 6 and the other is a 7 is ____________
- (2 pts) The number of card pairs such that they have the ranks $\{6,7\}$, $\{2,6\}$, or $\{A,2\}$ is ____________
- (3 pts) The size of the set of all possible pairs ,ignoring order (the size of the sample space) is ____________
- (3 pts) The conditional probability is ____________

Correct Answers:

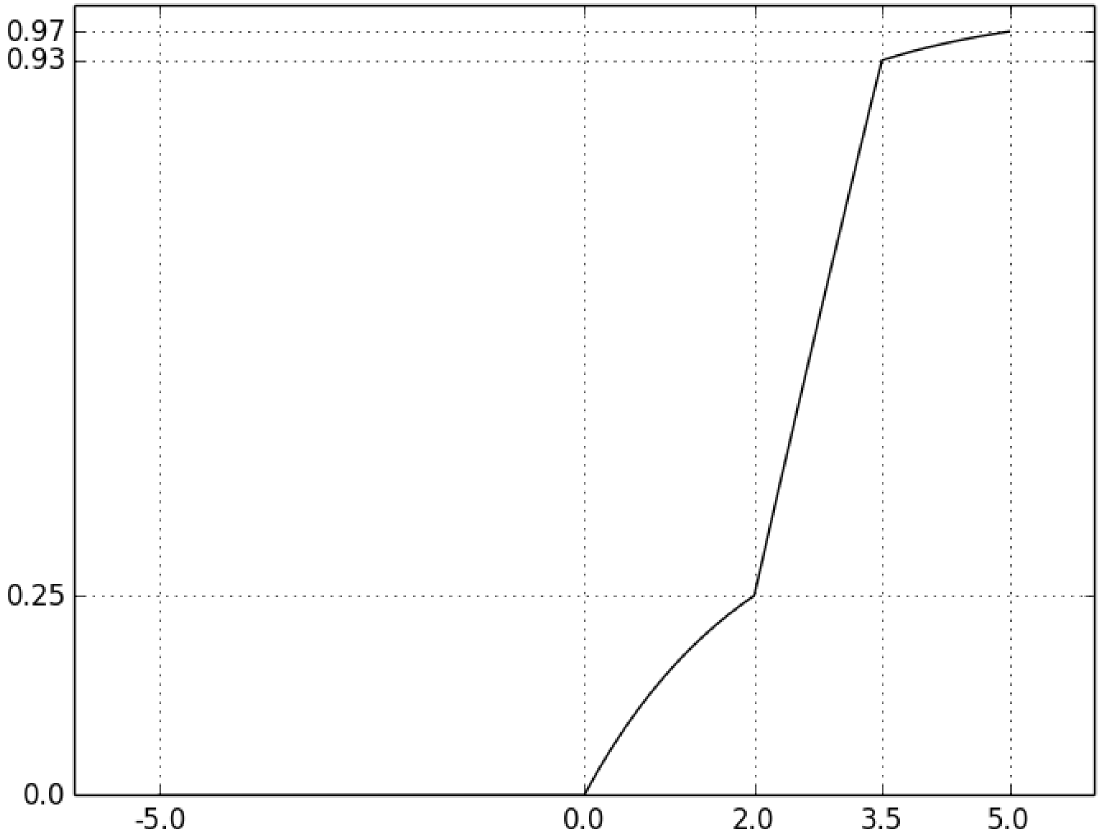
- `4^2`
- `3*4^2`
- `C(52-5,2)`
- `(3*4^2)/C(52-5,2)`

5. (10 pts)

Below is the CDF of a mixture distribution with **two** components.

One of the components is either a normal or an exponential distribution; the other is either a point mass or a uniform distribution.

Recall the CDF of exponential is $1-e^{-\lambda x}$.



Identify the component distributions:

- (3 pts) The exponential component has λ of 0.5, the probability of the exponential component assigns to the segment $[0,2]$ is ________________
- (3 pts) The weight of the exponential component is ________________
- (2 pts) The weight corresponding to the uniform component is therefore ________________
- (2 pts) The uniform component is over (________________ , ________________).

Correct Answers:

- `1-e^{-0.5*2}`
- `0.25/(1-e^{-0.5*2})`
- `1-0.25/(1-e^{-0.5*2})`
- `2`
- `3.5`

6. (9 pts)

An airline company is considering a new policy of booking as many as 227 persons on anairplane that can seat only 220.(Past studies have revealed that only 89% of the booked passengers actually arrive for the flight.)

We want to compute the probability that the number of passengers arriving for the flight is more than 220.

Denote by X_i the binary random variable that is 1 if passenger i arrived for the flight and 0 otherwise.

Denote by n=227 the number passengers that have booked the flight and by $S = \sum_{i=1}^{n} X_i$ the number of passengers that arrived for the flight.

- (3 pts) What is the mean of the number of passengers that arrive for the flight? $E(S) =$ __________________
- (3 pts) What is the standard deviation ? $\sqrt{var(S)} =$ __________________
- (3 pts) Using central-limit theorem and the Q function, estimate the probability that if the company books 227 persons, not enough seats will beavailable.

Correct Answers:

- `227*0.89`
- `sqrt (227 * 0.89 * (1 - 0.89))`
- `Q((220-(227*0.89))/(sqrt (227 * 0.89 * (1 - 0.89))))`

7. (9 pts)

Let X be a number that is uniformly distributed in $U(18, 96)$. Let S be the average of n = 34 samples of X: $S = \frac{1}{n}\sum_{i=1}^{n} X_i$.

1. (3 pts) $\mathbb{E}(S) =$ ____________________

2. (3 pts) $Var(S) =$ ____________________ **Hint:** The formular for the variance of the U(a, b) is $\frac{(b-a)^2}{12}$

3. (3 pts) Using Q function in your answer, what is the approximate probability of $P(S > 88)$ ____________________

Correct Answers:

- `(18+96)/2`
- `(96-18)^2/(12*34)`
- `Q((88-(18+96)/2)/(sqrt{(96-18)^2/(12*34)}))`

8. (9 pts)

Suppose $X_1, ..., X_{94}$ are IID random variables where $X_i \in \{0, 1\}$ and $P(X_i = 1) = 0.5$ for $i = 1, ..., n$.

Define a random variable $Y = \sum_{i=1}^{94} (-1)^{i+1} X_i = X_1 - X_2 + X_3 - X_4 + ... - X_{94}$

(3 pts) What is $E[Y]$? ____________________

(3 pts) What is $Var[Y]$? ____________________

(3 pts) Using Chebyshev's inequality, give an upper bound on $P(|Y| > 26)$? ____________________

Correct Answers:

- `0`
- `94*0.5*(1-0.5)`
- `94*0.5*(1 - 0.5)/26^2`

9. (20 pts)

Let X_1, X_2, X_3, X_4 be IID binary random variables for which $P(X_i = -1) = P(X_i = +1) = 1/2$.

- Define $Y_1 = X_1 + X_2$ and $Y_2 = X_3 + X_4$.

  (1 pt) Are Y_1, Y_2 independent (yes/no)? ____________________

  (2 pts) For any $i \in \{1,2,3,4\}$, the expectation $E(Y_i)$= ____________________ .

  Recall that for any two random variables A, B, the covariance is defined as $Cov(A,B) = E(AB) - E(A)E(B)$.

  (2 pts) For any $i \in \{1,2,3,4\}$, $E(X_i^2)$= ____________________ .

  (2 pts) For any $i \neq j$, $E(X_i X_j)$= ____________________ .

  (3 pts) $Cov(Y_1, Y_2)$= ____________________

- Define $Y_3 = X_2 + X_3$.

  (1 pt) Are Y_1, Y_3 independent (yes/no)? ____________________

  (1 pt) Are Y_2, Y_3 independent (yes/no)? ____________________

  (2 pts) $Cov(Y_1, Y_3)$= ____________________ .

- Recall the definition of the correlation coefficient:$Corr(A,B) = \frac{Cov(A,B)}{\sqrt{var(A)var(B)}}$.

  (2 pts) For any $i \in \{1,2,3\}$, $var(Y_i)$= ____________________ .

  (2 pts) $Corr(Y_1, Y_2)$= ____________________ .

  (2 pts) $Corr(Y_1, Y_3)$= ____________________ .

Correct Answers:

- 1
- 0
- 1
- 0
- 0
- 0

- 0
- 1
- 2
- 0
- 0.5

10. (11 pts)

Suppose you have a web server, with HTTP request packets arriving independently at random. The expected time between requests is 0.01.

(2 pts) What is the expected number of packets that arrive per second? ____________________

(3 pts) What is the probability that over 200 packets will arrive during a particular second? ____________________

(3 pts) Consider an hour as a sequence of one second bins (i.e. 3600 1-second bins). What is the expected number of bins during which more than 200 packets will arrive?

____

(3 pts) What is the probability that the time gap between two packets is larger than 0.03? ____________________

Note1: In one of the answers above you might need to use an infinite sum.

Note2: You may find the following formulas useful:

Poisson PDF: $P(X=k) = \frac{\lambda^k}{k!}e^{-\lambda}$

Poisson CDF: $P(X \leq k) = \sum_{i=0}^{k} \frac{\lambda^i}{i!}e^{-\lambda}$

Exponential PDF: $P(X=t) = \lambda e^{-\lambda t}$

Exponential CDF: $P(X \leq t) = 1 - e^{-\lambda t}$

Correct Answers:

- 100
- $1 - \sum_{i=0}^{200} \frac{100^i}{i!} e^{-100}$
- $3600 * (1 - \sum_{i=0}^{200} \frac{100^i}{i!} e^{-100})$
- $e^{-100*0.03}$

11. (16 pts)

You have an algorithm A for testing whether a Boolean formula f is satisfiable or not, but it is only correct with probability $2/3$. More precisely, you can repeatedly run A on the same formula f, and each time it outputs the correct answer with probability $2/3$.

To reduce the probability of error, you run A(f) n times, and return the majority answer. What should n be if you want the probability of error to be less than 0.05?

Let C_i be a binary random variable indicating whether the i^{th} execution of algorithm A is correct. Let $C = (C_1 + C_2 ... C_n)/n$.

- (2 pts) What is the minimum value of C such that our method of returning the majority answer will be correct? __________________
- (2 pts) What is $\mathbb{E}(C)$? __________________
- (2 pts) What is $var(C)$? __________________ (Use n as variable in this answer)

Approach 1:Chebyshev's inequality says for random variable Y with mean μ and for any positive number $a > 0$, $P(|Y - \mu| \geq a) \leq var(Y)/a^2$

- (3 pts) Using Chebyshev's inequality, what is an upper bound on the probability your "majority algorithm" is incorrect?

__________________ (Use n as variable in this answer)

- (2 pts) What is the lower bound on n so that the probability that the "majority algorithm" makes an error is at most 0.05?

____

Approach 2:Using Central Limit Theorem, approximate the distribution of C as a normal.

- (3 pts) What is the z-score of $C = 0.5$ __________________ (Use n as variable in this answer)
- (2 pts) What is the lower bound on n so that the "majority algorithm" makes an error with probability at most 0.05? __________________ (Use Q function in your answer)

Correct Answers:

- `1/2`
- `0.666667`
- `(1/n)*0.666667*(1-0.666667)`
- `(1/n)*0.666667*(1-0.666667)/((0.666667-0.5)**2)`
- `20*0.666667*(1-0.666667)/((0.666667-0.5)**2)`
- `(0.5-0.666667)/(sqrt(0.666667*(1-0.666667)/n))`
- `(Qinv(0.05)*sqrt(0.666667*(1-0.666667))/(0.5-0.666667))**2`