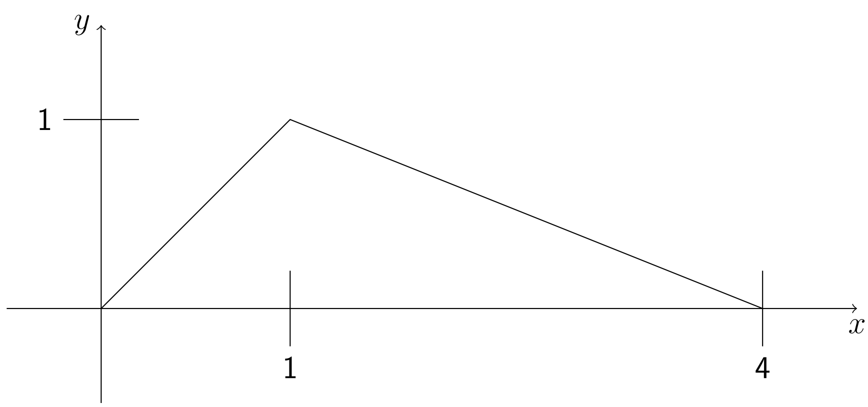
Math 3225 : Professor Lacey : Final Exam : December 2011

You Should Include Key Calculations for Each Problem! Just the Answers is NOT Enough.

Name: ______________________

Question:	1	2	3	4	5	6	7	8	Total
Points:	20	15	25	20	25	20	20	20	165
Score:									

1. Let (X, Y) be selected uniformly and at random in the triangle below.



   (a) (5 points) Find $\mathbb{E}(X \mid Y = 3/4)$.

   (b) (5 points) Describe the marginal distribution of Y.

   (c) (10 points) Calculate $\mathbb{E}Y^2$.

2. (15 points) A box contains 7 red and 8 white chips. Two chips are drawn, answer the probability that both

   (a) What is the chance that both chips are red?

   (b) What is the chance that second is red, if the first one is white?

   (c) What is the chance that both are red, given that at least one of the two chips is red?

3. (25 points) A bet on a split in Roulette has 1 chance in 19 of winning. You play this bet 38 times, and let X be the number of wins.

   (a) The exact distribution of X is

   (b) The exact probability $\mathbb{P}(X \geq 1) =$

   (c) The *approximate* distribution of X is

   (d) And the approximation gives $\mathbb{P}(X \geq 1) \simeq$

4. (20 points) Let X be the winning from a bet on a split in roulette, and let Y be the winnings from a bet on red. The distribution of X and Y is given below.

$$X = \begin{cases} 17 & \text{2 chance in 38} \\ -1 & \text{36 chances in 38} \end{cases} \qquad Y = \begin{cases} 1 & \text{18 chances in 38} \\ -1 & \text{20 chances in 38} \end{cases}$$

Both X and Y have expected value of $-1/19$.

(a) Calculate the variance of X and Y.

(b) If you plan on placing a modest number of bets on roulette, and you would like to maximize your chance of coming out ahead. Which bet should you make? A single number or a bet on red? Why?

(c) A survey of medical studies on hepatitis finds higher mortality rates in those studies with smaller sample sizes. Give a purely probabilistic explanation of this.

5. (25 points) (Poisson Approximation!) Let 150 randomly selected people be labeled by $1 \leq i \leq 100$. Set

$$X_{i,j,k} = \begin{cases} 1 & \text{People } i, j, k \text{ all share a birth date} \\ 0 & \text{otherwise} \end{cases}$$

(a) For $i < j < k$, $P(X_{i,j,k} = 1)$ equals what?

(b) Are these random variables independent?

(c) $Y = \sum_{1 \leq i < j < k \leq 150} X_{i,j,k}$ is approximately Poisson with $\lambda =$??

(d) Estimate the chance that no three people among the 150 share a birth date.

6. (20 points) In 2005[1], 29,000 children, up to 15 years of age, who were diagnosed with cancer between the years of 1962 to 1995, living in England and Wales.

   They were able to map how far each child lived from a high voltage overhead power line. Comparing the children who had cancer with a control group of 29,000 children without cancer but who lived in comparable districts, found that children whose birth address was within 200 meters of an overhead power line had a 70% increased risk of leukemia.

   The study concludes that since "there is no biological mechanism to explain the higher risk", the results, "although statistically significant, may be due to chance".

   "To put these results in perspective, our study shows that about five of the 400 cases of childhood leukemia every year may be linked to power lines - which is about 1% of cases," says Gerald Draper at Oxford University, who led the study. "The condition is very rare and people living near power lines should have no cause for concern."

   (a) Is there an underlying model for the data that the study has in mind? What is it? Answer in terms of the two study groups.

   (b) Does the phrase of 'statistically significant' make sense in the context of the data?

   (c) Comment on the level of statistical significance that the study found, and the nature of conclusion.

   (d) Comment on the role of leukemia in the study. How was it incorporated into the study? How did the researchers come to the conclusion about childhood leukemia?

[1] British Medical Journal (vol 330, p 1290)

7. (20 points) A rare type of genetic mutation in makes the bacterium *E. coli* antibiotic resistant. A sample of $n = 150$ petri dishes containing an antibacterial agar were plated with 10^6 bacteria. After a period of time, the number of colonies on each dish were counted. Assume that each colony formed from a single resistant bacterium.

num. of colonies	0	1	2	3	4
observed frequency	85	52	9	3	1

(a) The number of colonies should be Poisson in distribution. Why?

(b) Test the hypothesis that the distribution is Poisson with $\lambda = \overline{x} = 0.5$. Include:

i. a description of the Null Hypothesis,

ii. a table of expected frequencies. What should you do with the one dish with four colonies?

iii. the χ^2 sum, with the degrees of freedom and the conclusion of the test.

8. (20 points) A random sample of 50 women is classified according to age and cholesterol level, generating the table below.

	Cholesterol Level			
Age	< 180	180—210	> 210	Totals
< 50yrs	5	11	9	25
≥ 50yrs	5	7	13	25
Totals	10	18	22	50

(a) Test the null hypothesis of the independence of age and cholesterol levels.

(b) Give a table of expected occurrences if the null hypothesis is true.

(c) Give the value of the χ^2 statistic, and its degrees of freedom.

(d) Give the result of the test.