 Name ____________________

Please place your name on all pages NOW

IE 202 – Introduction to Probability
Final Exam

Note that the last page is a table of Normal distribution values like Table A.3 in the text in case you need it.

Problem 1: (20 Percent – 5% each part)

A CCD is composed of millions of pixels that are used to record pictures in most digital cameras. For example, the CCD on the Nikon D200 has 10.2 million pixels. Sometimes a pixel will be defective.

The probability that a pixel will be defective is $1x10^{-7}$. Assume that whether or not any individual pixel works is *independent* of the probability of another pixel on the same screen working.

a) Find the probability that a screen will have NO defective pixels on a CCD in a Nikon D200.

$$\lambda = 10^{-7} \bullet 10.2 \bullet 10^{6} = 1.02$$
$$P(\text{0 defects}) = e^{-1.02} = 0.3606$$

b) Find the probability that a screen will have exactly ONE defective pixel.

$$\lambda = 10^{-7} \bullet 10.2 \bullet 10^{6} = 1.02$$
$$P(\text{1 defect}) = 1.02e^{-1.02} = 0.3678$$

c) A camera is deemed defective if FOUR or more pixels on the CCD are defective. What is the probability that a camera will be defective?

$$\lambda = 10^{-7} \bullet 10.2 \bullet 10^{6} = 1.02$$
$$\begin{aligned} P(\text{3 defects or fewer}) &= e^{-1.02} + 1.02e^{-1.02} + \frac{1.02^2 e^{-1.02}}{2!} + \frac{1.02^3 e^{-1.02}}{3!} \\ &= 0.3606 + 0.3678 + 0.1876 + 0.0638 \\ &= 0.9798 \end{aligned}$$
$$P(\text{DEFECTIVE CAMERA}) = 0.0202$$

d) Each month Nikon produces 500 of these cameras. What is the (approximate) probability that the number of defective cameras in a month is more than 15?

Approximate the Binomial by a Normal with

$\mu = 500(0.020239) = 10.12$

$\sigma^2 = 500(0.020239)(1-0.020239) = 9.9147$

$\sigma = 3.1488$

$P(\text{more than 15}) = 1 - P(15 \text{ or less}) = 1 - P\left(Z \le \frac{15.5-10.12}{3.1488}\right)$

$= 1 - P(Z \le 1.71) = 1 - 0.956 = 0.044$

Problem 2: (40 percent – 4% each part)

Consider the following random variable: $Y = a + X$, where a is a constant and X is itself a random variable that has an Exponential distribution with parameter λ.

a) What is the mean of Y in terms of λ and a?

$$E(Y) = a + \frac{1}{\lambda}$$

b) What is the variance of Y in terms of λ and a?

$$Var(Y) = \frac{1}{\lambda^2}$$

c) What is the probability density function of Y in terms of λ and a?

$$f_Y(y) = \begin{cases} 0 & y \le a \\ \lambda e^{-\lambda(y-a)} & y > a \end{cases}$$

d) What is the cumulative distribution of Y in terms of λ and a?

$$F_Y(y) = \begin{cases} 0 & y \le a \\ 1 - e^{-\lambda(y-a)} & y > a \end{cases}$$

e) The Exponential distribution has the memoryless property. Does Y have the memoryless property? **Justify your answer** One way to do so would be in terms of the probability density function of Y, given that the random variable Y takes on a specific value y_0 that is strictly greater than a?

$$P\left(Y > z \middle| Y > y_0\right) = \frac{P\left(Y > z \text{ AND } Y > y_0\right)}{P\left(Y > y_0\right)} = \frac{e^{-\lambda(z-a)}}{e^{-\lambda(y_0-a)}} = e^{-\lambda(z-y_0)}$$

So, it is not memoryless

f) Now suppose that we have a third random variable Z which is the sum of 3 independent identically distributed random variables each with the same distribution as Y. In other words, $Z = Y_1 + Y_2 + Y_3$, where each Y_i has the same distribution as Y above. What is the mean of Z in terms of a and λ? *Note that you should be able to answer this even if you could not answer parts (c, d, or e).*

$$E(Z) = 3a + \frac{3}{\lambda}$$

g) What is the variance of Z in terms of a and λ?

$$Var(Z) = \frac{3}{\lambda^2}$$

h) In words, how would you describe the distribution of Z?

This is an Erlang-3 distribution shifted 15 units to the right.

i) Suppose λ=0.1 and $a = 5$, what are the mean and variance of Z?

$$E(Z) = 3a + \frac{3}{\lambda} = 15 + 30 = 45$$

$$Var(Z) = \frac{3}{\lambda^2} = 300$$

j) Suppose λ=0.1 and $a = 5$, what is the probability that Z is less than or equal to 60?

We know that Z must be at least 15 and that the rest of the distribution is an Erlang-3 distribution. Thus, this is equivalent to the Probability that an Erlang-3 distribution is less than or equal to 45, which is the probability of 3 or more exponential events with rate λ=0.1 in 45 minutes, which is the 1 minus the probability of 2 or fewer events in this time. Thus we want

$$1 - \left(e^{-4.5} + \frac{4.5e^{-4.5}}{1!} + \frac{4.5^2 e^{-4.5}}{2!} \right) = 1 - 0.1736 = 0.8364$$

Problem 3: (20 percent – 4% each part)

Passengers arrive at a bus stop at random. In other words, their arrival times are independent of one another and independent of the bus arrival times. Bus headways (the time between busses) follow the distribution below:

Headway (min)	5	10	20
Probability	0.2	0.2	0.6

a) Find the mean headway.

$$E(H) = 0.2(5) + 0.2(10) + 0.6(20) = 1 + 2 + 12 = 15$$

b) Find the variance of the headway.

$$E(H^2) = 0.2(25) + 0.2(100) + 0.6(400) = 5 + 20 + 240 = 265$$

$$Var(H) = E(H^2) - E^2(H) = 265 - 15^2 = 265 - 225 = 40$$

c) Find the mean waiting time of a randomly selected passenger. *Note that more people will arrive during the longer headways and so you are more likely to "see" them when you select a passenger at random.*

Headway	5	10	20
Probability	0.2	0.2	0.6
H*prob	1	2	12
H*H*prob	5	20	240
Waiting time	2.5	5	10
People*time	12.5	50	200
People	5	10	20

E(People*Time)	132.5
E(People)	15
Mean waiting time	8.833

So mean waiting time is 8.833 minutes.

d) Now suppose that the headway actually has a distribution given by $H = a + X$, where a is a constant and X is itself a random variable that has an Exponential distribution with parameter λ *(Note that this is the same distribution that we had in problem 2, parts a and b, above.)* Recall that the average waiting time is given by $\frac{E(H)}{2} + \frac{Var(H)}{2E(H)}$. In terms of a and λ what is the average waiting time of a passenger if this is the distribution?

$$E(H) = a + \frac{1}{\lambda} = \frac{a\lambda + 1}{\lambda}$$

$$Var(H) = \frac{1}{\lambda^2}$$

$$\begin{aligned} E(wait) &= \frac{E(H)}{2} + \frac{Var(H)}{2E(H)} \\ &= \frac{a\lambda+1}{2\lambda} + \frac{1}{2\lambda^2}\frac{\lambda}{a\lambda+1} \\ &= \frac{a}{2} + \frac{1}{2\lambda} + \frac{1}{2\lambda(a\lambda+1)} \end{aligned}$$

e) In part (d), if the parameters of the headway distribution (measured in seconds) are given by $a = 60$ seconds and $\lambda = 1/540$, what is the average waiting time of a passenger?

$$\begin{aligned} E(wait) &= \frac{E(H)}{2} + \frac{Var(H)}{2E(H)} \\ &= \frac{a\lambda+1}{2\lambda} + \frac{1}{2\lambda^2}\frac{\lambda}{a\lambda+1} \\ &= \frac{a}{2} + \frac{1}{2\lambda} + \frac{1}{2\lambda(a\lambda+1)} \\ &= \frac{60}{2} + \frac{1}{2(1/540)} + \frac{1}{2(1/540)(60(1/540)+1)} \\ &= 30 + 270 + \frac{540}{2((1/9)+1)} = 30 + 270 + \frac{270(9)}{10} = 543 \end{aligned}$$

or a little more than 9 minutes, when the average time between busses is 5 minutes.

Problem 4: (24 percent – 4% each part)

The following data are from
http://apps.collegeboard.com/search/compare_schools.jsp?

SCHOOL	SAT-V Mid 50%	SAT – M Mid 50%
Northwestern University (Evanston, IL)	650 to 740	670 to 760
Brown University (Providence, RI)	650 to 760	660 to 760

Assume that the scores at each school are Normally distributed with the values shown above. In other words, the Verbal SAT Scores at Northwestern are Normally distributed with the middle 50% falling between 650 and 740.

a) For Northwestern University, find the mean and standard deviation of the Verbal SAT scores?

The mean is 695
0.675 standard deviations is equal to 45, so 1 standard deviation is 66 2/3. Note that you get this by observing that you have to go out 0.675 standard deviations from the mean to get to a probability of 0.25.

b) Find the mean and standard deviation of the Math SAT scores at Northwestern University.

The mean is 715
0.675 standard deviations is equal to45, so 1 standard deviation is 66 2/3.

c) Find the mean and standard deviation of the verbal and math SAT scores at Brown University.

Verbal: Mean = 705, standard deviation = 81.48
Math: Mean = 710, standard deviation = 74.07.

d) What is the distribution of the total (math plus verbal) SAT score at Northwestern University, assuming that the math and verbal scores for any student are independent? **Give the name of the distribution and all relevant parameter values.**

Normal with a mean of 1410 and a standard deviation of 94.28.

e) What is the distribution of the total (math plus verbal) SAT score at Brown University, assuming that the math and verbal scores for any student are independent? **Give the name of the distribution and all relevant parameter values.**

Normal with a mean of 1415 and a standard deviation of 110.12.

f) Suppose 36 students from Northwestern are sampled at random and 49 students from Brown University are sampled at random. For each student their total (math plus verbal) SAT Score is recorded. The average score at NU is then computed as is the average of the 49 scores at Wash U. What is the probability that the average for the 36 NU students will be **more** than the average of the 49 Wash U students? *Hint: First write down the distribution of the difference between the two averages.*

Normal with a mean of -5 and a standard deviation of 22.23. So we want the probability that a standard normal random variable takes on a value of 5/22.23 = 0.225 or more. This probability is 1-0.5890=0.4110.

Problem 5 (16 percent – 3% each part except part (e) which is 4%):

You have collected data on the number of e-mail messages that you get every day. The number has a Poisson distribution with a rate of $\lambda = 15$ per day. Of these, you find that 70% are typically spam!

a) Given that you get 20 e-mails one day, what is the probability that exactly 16 of these e-mails are spam?

This is a Binomial distribution with n=20 and p=0.7

$$P(16 \text{ spams}) = \binom{20}{16} 0.7^{16} 0.3^{4} = 0.1304$$

b) What is the probability of getting 20 e-mails in one day?

This is a Poisson distribution with λ=15

$$P(20 \text{ e-mails}) = \frac{15^{20} e^{-15}}{20!} = 0.0418$$

c) What is the joint probability of getting 20 e-mails of which 16 are spam?

$$P(20 \text{ e-mails}) P(16 \text{ of } 20 \text{ are spam}) = \left[\frac{15^{20} e^{-15}}{20!}\right] \bullet \left[\binom{20}{16} 0.7^{16} 0.3^{4}\right] = 0.00545$$

d) What are the mean and variance of the number of spam e-mails, given that you get 20 e-mails in a particular day?

$$E(\# \text{ spam}) = 20(0.7) = 14$$

$$Var(\# \text{ spam}) = 20(0.7)(0.3) = 4.2$$

e) Are the number of e-mails you receive and the number that are spam independent?

No. Clearly, if you know that you received 16 spam e-mails in 1 day, this tells you that you received at least 16 e-mails that day. Thus, they cannot be independent.