1. The weight of the eggs produced by a certain breed of hen is Normally distributed with mean 70 grams(g) and standard deviation 5 g.

a) Sketch the graph of the normal distribution that corresponds to the weight of the eggs.



b) What is the probability that a randomly selected egg is more than 62 grams?



$$P(X > 62) = P\left(Z > \frac{62 - 70}{5}\right) = P(Z > -1.6)$$

Method 1
= P(Z < 1.6) = 0.9452
Method 2
= 1 - P(Z < -1.6) = 1 - 0.0548 = 0.9452

c) What is the probability that a randomly selected egg is between 62 and 69.4 grams?

$$P(62 < X < 69) = P\left(\frac{62 - 70}{5} < Z < \frac{69.4 - 70}{5}\right) = P(-1.6 < Z < -0.12) = 0.4522 - 0.0548$$
$$= 0.3974$$

d) How much does an egg need to weigh in order for it to be the largest 5% of weights of all of the eggs?



The z-score corresponding to the top 5% is z = 1.645OR P(Z > z) = 0.05 ==> P(Z < z) = 1 - 0.05 = 0.95 ==> z = 1.645

x = 70 + 1.645*5 = 78.225An egg needs to weight at least 78.225 grams in order for it to be the largest 5%.

2. The editor of a statistics text would like to plan for the next edition. A key variable is the number of pages that will be in the final version. Text files are prepared by the authors using a word processor called LaTeX, and separate files contain figures and tables. For the previous edition of the text, the number of pages in the LaTeX files can easily be determined, as well as the number of pages in the final version of the text. Here are the data:

	Chapter												
	1	2	3	4	5	6	7	8	9	10	11	12	13
LaTeX pages	77	73	59	80	45	66	81	45	47	43	31	46	26
Text pages	99	89	61	82	47	68	87	45	53	50	36	52	19

a) Find the mean and the standard deviation for LaTex pages and Text pages.

	LaTeX	Text
summations from Excel	$\sum x = 719, \sum x^2 = 43917$	$\sum x = 788, \sum x^2 = 54264$
$\bar{x} = \frac{\sum x}{n}$	$\frac{719}{13} = 55.31$	$\frac{788}{13} = 60.62$
$s = \sqrt{\frac{\sum x^2 - \frac{(\sum s)^2}{n}}{n-1}}$	$\sqrt{\frac{43917 - \frac{719^2}{13}}{12}} = 18.60$	$\sqrt{\frac{54264 - \frac{788^2}{13}}{12}} = 23.27$

Solution to STAT 350 Exam 1 Review Problems

Variable	Mean	StDev
LaTeXPages	55.31	18.60
TextPages	60.62	23.27

b) Find the five number summaries for LaTex pages and Text pages.

Sorted data

	1	2	3	4	5	6	7	8	9	10	11	12	13
LaTex	26	31	43	45	45	46	47	59	66	73	77	80	81
Text	19	36	45	47	50	52	53	61	68	82	87	89	99
				Q ₁			Ĩ			Q ₃			

$$d_{1} = \frac{n}{4} = \frac{13}{4} = 3.25 \implies 4$$
$$d_{3} = \frac{3n}{4} = \frac{(3)(13)}{4} = 9.75 \implies 10$$

Variable	Minimum	Q1	Median	Q3	Maximum
LaTeXPages	26.00	45.00	47.00	73.00	81.00
TextPages	19.00	47.00	53.00	82.00	99.00

c) Draw a side-by-side boxplot for LaTex pages and Text pages.



3, Suppose that 8% of tires manufactured by a certain company are defective. Assume that we have a random sample of 16 of these tires (enough for 4 cars). Note: This is a binomial distribution. What is the probability that at least one of the tires are defective? Write down the complete formula that is to be used.

n = 16, p = 0.08

$$P(X \ge 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - {\binom{16}{0}} 0.08^{0} (1 - 0.08)^{16} = 1 - 0.92^{16} = 1 - 0.263$$

$$= 0.737$$

4. Cars pass your house at an average of 2.5 cars per day (you live in the country). Let X which is the number of cars that pass by your house have a Poisson distribution.

a) What is the probability that the at least 2 cars pass by your house on a given day?

 $P(X \ge 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1) = 1 - \frac{e^{-2.5} \cdot 2.5^0}{0!} - \frac{e^{-2.5} \cdot 2.5^1}{1!}$ = 1 - 0.082 - 0.205 = 0.713

b) In the next 2 months (60 days), what is the probability that the <u>average</u> number of cars that passes your house per day is at least 2?

Since it says 'average', this is a sampling distribution.

 $\mu_{\bar{x}} = \lambda = 2.5$ $\sigma_{\bar{X}} = \frac{\sqrt{\lambda}}{\sqrt{n}} = \frac{\sqrt{2.5}}{\sqrt{60}} = 0.204$ $P(\bar{X} \ge 2) = P\left(Z \ge \frac{2 - 2.5}{0.204}\right) = P(Z \ge -2.45) = 1 - P(Z < -2.45) = 1 - 0.0071 = 0.9929$

c) If the number of cars that passes your friend's house, Y, has a Poisson distribution with average of 25 per day, what is the standard deviation of Y - X?

$$\sigma_{Y-X} = \sqrt{\sigma_Y^2 + \sigma_X^2} = \sqrt{25 + 2.5} = 5.244$$

d) If the vehicles that pass by your house are either trucks (T) or cars (C), what is the sample space of the next 2 vehicles that pass by your house?

 $S = \{TT, TC, CT, CC\}$

e) Suppose that 55% of the vehicles are trucks, 25% of the vehicles are red and 17% are red trucks. In addition, the vehicles that are not trucks are cars (we are assuming that there are no motorcycles). Given that a vehicle is red, what is the probability that it is a truck?

 $P(T) = 0.55 \quad P(R) = 0.25 \quad P(T \text{ and } R) = 0.17 \qquad P(C) = 0.45$ $P(T|R) = \frac{P(T \text{ and } R)}{P(R)} = \frac{0.17}{0.25} = 0.68$

Solution to STAT 350 Exam 1 Review Problems

5. According to a book review published in the Wall Street Journal on Sep 26, 2012, 1% of 40year-old women have breast cancer. 80% of these women who actually have breast cancer will have a positive mammogram. 10% of 40-year-old women who do not have breast cancer will also have a positive mammogram. If a 40-year-old woman has positive mammogram, what is the probability that she has breast cancer?

Solution: We want to find P(Cancer | Test +). [a tree diagram will be helpful] We know: $P(Cancer)=0.01; \Rightarrow P (No Cancer) = 0.99$ P(Test + | Cancer) = 0.8P(Test + | NO Cancer) = 0.1

It is acceptable to draw a tree diagram to solve the problem.

So, P(Test +) = P (Test + and Cancer) + P(Test + and NoCancer) =0.01*0.8 + 0.99*0.1 = 0.107 P(Cancer | Test +) = P(Cancer and Test +) /P(Cancer) = 0.01*0.8/0.107 = 0.075

- 6. The credit manager for a local department store discovers that 88% of all the store's credit card holders who defaulted on their payments were late (by a week or more) with two or more of their monthly payments before failing to pay entirely (defaulting). This prompts the manager to suggest that future credit be denied to any customer who is late with two monthly payments. Further study shows that 3% of all credit customers default on their payments and 40% of those who have not defaulted have had at least two late monthly payments in the past.
 - a) What is the probability that a customer who has two or more late payments will default?

 $\begin{array}{ll} \mathsf{P}(\mathsf{Late}|\mathsf{Default}) = 0.88 & \mathsf{P}(\mathsf{Default}) = 0.03 & \mathsf{P}(\mathsf{Late}|\mathsf{Default^c}) = 0.4 \\ \mathsf{P}(\mathsf{Default^c}) = 1 - \mathsf{P}(\mathsf{Default}) = 1 - 0.03 = 0.97 \end{array}$

It is acceptable to draw a tree diagram to solve the problem.

$$\begin{split} P(Default|Late) &= \frac{P(Late|Default)P(Default)}{P(Late|Default)P(Defaul) + P(Late|Default^c)P(Default^c)} \\ &= \frac{(0.88)(0.03)}{(0.88)(0.03) + (0.4)(0.97)} = 0.0637 \end{split}$$

b) Under the credit manager's policy, in a group of 100 customers who have their future credit denied, how many would we expect *not* to default on their payments?

P(Default) = 0.03 $P(Default^{c}) = 0.97$ P(Late|Default) = 0.88 $P(Late|Default^{c}) = 0.4$

It is acceptable to draw a tree diagram to solve the problem.

$$P(Default^{c}|Late) = \frac{P(Late|Default^{c})P(Default^{c})}{P(Late|Default^{c})P(Default^{c}) + P(Late|Default)P(Default)}$$
$$= \frac{(0.4)(0.97)}{(0.4)(0.97) + (0.88)(0.03)} = 0.936$$
$$(0.936)(100) = 93.6 \text{ therefore between } 93 \text{ and } 94 \text{ would NOT default.}$$

c) Does the credit manager's policy seem reasonable? Explain your response

No, the policy is not reasonable.

Only 3% of the customers default. Of those who are late, only 6.37% default. Knowing that a customer is late on payments does not dramatically increase the chance that they will default on the payment though it does make the probability higher.

7. Servings of fruits and vegetables. The following table gives the distribution of the number of servings of fruits and vegetables consumed per day in a population.

Number of servings X	0	1	2	3	4	5
Probability	0.3	0.1	0.1	0.2	0.2	0.1

Find the mean and the standard deviation for this random variable.

$$\mu = \sum_{i=0}^{5} x_i * p_i = 0 * 0.3 + 1 * 0.1 + 2 * 0.1 + \dots + 5 * 0.1 = 2.2$$

$$\sigma^2 = \left[\sum_{i=0}^{5} x_i^2 * p_i\right] - (\mu^2) = [0 * 0.3 + 1 * 0.1 + 4 * 0.1 + 9 * 0.2 + 16 * 0.2 + 25 * 0.1] - 2.2^2$$

$$\sigma = \sqrt{3.16} = 1.78$$

8. f(x) = k x (2 - x), 0 < x < 2

5

a) What is the constant k that makes the above function a valid density function?

$$\int_{0}^{2} kx(2-x)dx = k \int_{0}^{2} (2x-x^{2})dx = k \left[\frac{2x^{2}}{2} - \frac{x^{3}}{3}\right]_{0}^{2} = k \left(4 - \frac{8}{3}\right) = \frac{4}{3}k = 1 \Rightarrow k = \frac{3}{4}$$

b) Find P(0< X <1.5)

$$P(0 < X < 1.5) = \int_0^{1.5} \frac{3}{4} x(2-x) dx = \frac{3}{4} \int_0^{1.5} (2x-x^2) dx = \frac{3}{4} \left[\frac{2x^2}{2} - \frac{x^3}{3} \right]_0^{1.5} = \frac{3}{4} (2.25 - 1.125)$$

= 0.84

c) Find the mean and the standard deviation for X.

$$E(X) = \int_0^2 x \frac{3}{4} x(2-x) dx = \frac{3}{4} \int_0^2 (2x^2 - x^3) dx = \frac{3}{4} \left[\frac{2x^3}{3} - \frac{x^4}{4} \right]_0^2 = \frac{3}{4} \left(\frac{16}{3} - 4 \right) = 1$$

$$E(X^{2}) = \int_{0}^{2} x^{2} \frac{3}{4} x(2-x) dx = \frac{3}{4} \int_{0}^{2} (2x^{3} - x^{4}) dx = \frac{3}{4} \left[\frac{2x^{4}}{4} - \frac{x^{5}}{5} \right]_{0}^{2} = \frac{3}{4} (8 - 6.4) = 1.2$$

$$\sigma^{2} = E(X^{2}) - (E(X))^{2} = 1.2 - 1^{2} = 0.2$$

$$\sigma = \sqrt{\sigma^{2}} = \sqrt{0.2} = 0.447$$

d) Find the 80th percentile of X. (i.e. find x so that P(X < x) = 0.8)

$$0.8 = \int_0^y \frac{3}{4}x(2-x)dx = \frac{3}{4}\int_0^y (2x-x^2)dx = \frac{3}{4}\left[\frac{2x^2}{2} - \frac{x^3}{3}\right]_0^y = \frac{3}{4}\left(y^2 - \frac{y^3}{3}\right) = 0.75y^2 - 0.25y^3$$

The integration finds the CDF.

Therefore, the equation that you need to find the roots for is $0.75y^2 - 0.25y^3 - 0.8 = 0$ I prefer to remove the fractions by multiplying the equation by 4 (this is not required) to obtain $3y^2 - y^3 - 3.2 = 0$

This is a cubic (this will NOT be on the exam) so you can solve it via a graphing calculator or Wolfram alpha (which is what I use)

The three roots are -0.91, 1.43, 2.48. The only one that is in the range of (0,2) is 1.43 so this is the answer.

9. $f(x) = k(e^{-x} + e^{-4x})$ for x > 0

a) What is the constant k that makes the above function a valid density function?

$$\int_0^\infty k(e^{-x} + e^{-4x})dx = k\left(\int_0^\infty e^{-x}dx + \int_0^\infty e^{-4x}dx\right) = k\left(\left[-e^{-x}\right]_0^\infty + \left[-\frac{1}{4}e^{-4x}\right]_0^\infty\right)$$
$$= k\left[-(0-1) - \frac{1}{4}(0-1)\right] = \frac{5}{4}k \Rightarrow k = \frac{4}{5}$$

b) Find P(X > 2).

$$P(X > 2) = \int_{2}^{\infty} \frac{4}{5} (e^{-x} + e^{-4x}) dx = \frac{4}{5} \left(\int_{2}^{\infty} e^{-x} dx + \int_{2}^{\infty} e^{-4x} dx \right) = \frac{4}{5} \left(\left[-e^{-x} \right]_{2}^{\infty} + \left[-\frac{1}{4} e^{-4x} \right]_{2}^{\infty} \right)$$
$$= \frac{4}{5} \left[-(0 - e^{-2}) - \frac{1}{4} (0 - e^{-8}) \right] = \frac{4}{5} (0.135) = 0.11$$

c) Find the mean and the standard deviation for X.

You could do this problem from scratch but that would involve integrating by parts. So I am using a different method.

I am recognizing that this is a linear combination of two independent exponential distributions, X has $\lambda = 1$ and Y has $\lambda = 4$. Therefore, Z = (0.8)[X + 0.25Y]

$$E(Z) = (0.8)[E(X) + 0.25E(Y)] = (0.8)\left[\frac{1}{1} + \frac{1}{4}\frac{1}{4}\right] = (0.8)[1.0625] = 0.85$$

 $Var(Z) = (0.8^{2})[Var(X) + 0.25^{2}Var(Y)] = (0.8^{2})\left[\frac{1}{1^{2}} + \frac{1}{4^{2}}\frac{1}{4^{2}}\right] = (0.8^{2})[1.0039] = 0.6425$ $\sigma = \sqrt{Var(Z)} = \sqrt{0.6452} = 0.8016$ **10.** Suppose the time until the earthquake in Lonely Mountain has an exponential density function with an average of 2 years.

Since this is an exponential distribution with E(X) = 2, $\lambda = \frac{1}{2} = 0.5$ The distribution is 0.5 e^{-x/2}.

a) Find the probability that the next earthquake happens within two years.

 $P(X < 2) = 0.5 \int_{0}^{2} e^{-x/2} dx = \left[-e^{-x/2}\right]_{0}^{2} = (-e^{-1} + 1) = 0.632$ or you can use the CDF F(X) = 1 - e⁻¹ = 0.632

b) Find the median time until the next earthquake in Lonely Mountain.

$$0.5 = 0.5 \int_{0}^{y} e^{-x/2} dx = \left[-e^{-x/2}\right]_{0}^{y} = \left(-e^{-y/2} + 1\right) \Rightarrow e^{-\frac{y}{2}} = 0.5$$

Again, you can use this problem by using the CDF.
Take the ln (natural log) of each side
 $-\frac{y}{2} = \ln(0.5) \Rightarrow y = -2\ln(0.5) = 2ln2 = 1.386$