# FALL 2006 FINAL EXAM SOLUTIONS
# MATH 218, DECEMBER 13, 2006

This is a solution set for the Math 218 final common exam held from 2:00 to 4:00 PM on Wednesday, December 13, 2006. There were ten problems, weighted equally (for a total of 200 points). These solutions are rather more full and complete than anything we could have expected from any student.

**Problem 1** (20 pts). 30 *marbles are sitting on a table:* 6 *red,* 9 *blue, and* 15 *green. All of the red marbles,* 3 *of the blue marbles, and* 10 *of the green marbles are covered in chocolate, because a child played with them while eating a piece of pie. The remaining marbles were out of the child's reach and remained clean.*

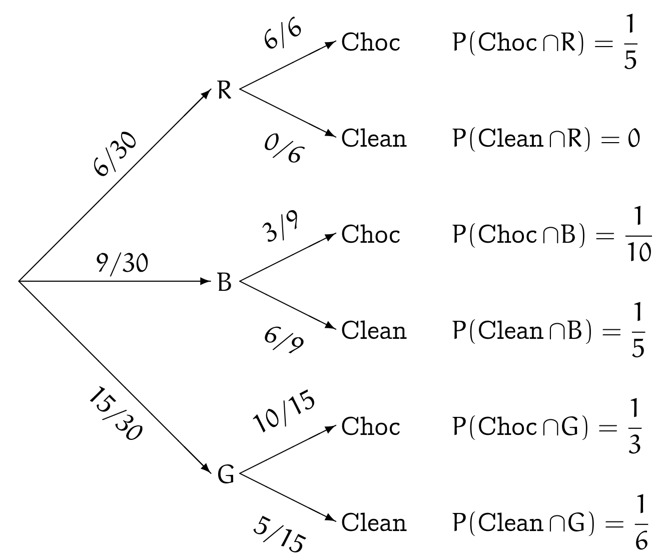
(a) *Draw a probability tree describing the situation. Be sure to include the labels of events, probabilities, and conditional probabilities.*
(b) *If a randomly selected marble is covered in chocolate, what is the probability that it is red?*
(c) *If a randomly selected marble is red, what is the probability that it is covered in chocolate?*
(d) *Are the events "red" and "covered in chocolate" independent?*

*Solution.*

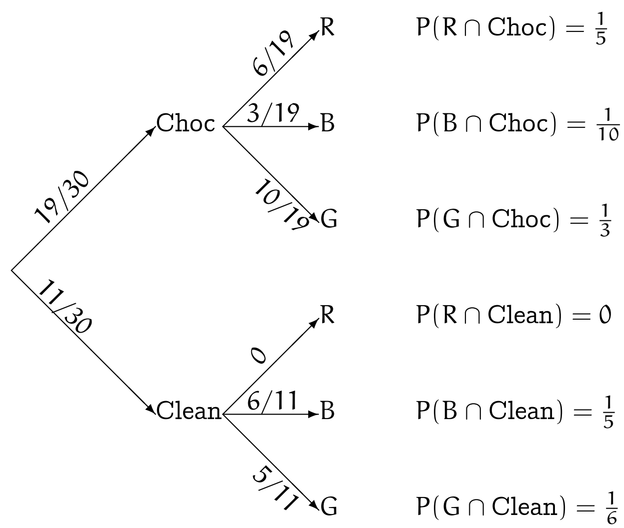
(a) The marbles are partitioned in two ways: according to color, and according to whether they're covered in chocolate or are clean. There are, accordingly, two ways you can draw the tree: in the first, the first level of the tree is based on color, and the second level based on chocolate/clean; in the second, the first level of the tree is based on chocolate/clean, and the second is based on color. The two ways are about equally easy.

Our first tree uses the color-first approach. Of course, you were instructed to reduce fractions to lowest terms, so instead of "10/15", for example, as a conditional probability in the tree, you should have written "2/3". We have left the fractions unreduced to make it more obvious where they came from. Also, the calculation of the $\cap$ probabilities at the right edge was not, strictly speaking, asked for; but they're too useful for the rest of the problem, and since they're trivial to calculate, you should probably always write them down.

Here's the first tree:




In the second possible tree description, we have to count the number of chocolate-covered marbles and the number of clean marbles. Obviously there are 19 chocolate-covered marbles (6 red, 3 blue, and 10 green), leaving 11 clean marbles (0 red, 6 blue, and 5 green). This tree has the advantage that all the probabilities are already in lowest terms:




(b)

$$P(R|\,Choc) = \frac{P(R \cap Choc)}{P(Choc)} = \frac{P(R \cap Choc)}{P(R \cap Choc) + P(B \cap Choc) + P(G \cap Choc)},$$

which from the tree in part (a) is

$$P(R|\,Choc) = \frac{1/5}{1/5 + 1/10 + 1/3} = \frac{6}{19} \approx 0.316.$$

Of course, this is even easier if you use the second tree: the conditional probability 6/19 is already built into the tree.

(c) You can do this with no computation, just a little thinking: you're *told* that *all* of the red marbles were covered in chocolate, so if you select a red marble, well. . . there's a 100% chance it's covered in chocolate. . .

Of course, you can also do the computation:

$$P(Choc\,|R) = \frac{P(Choc \cap R)}{P(R)} = \frac{P(Choc \cap R)}{P(Choc \cap R) + P(Clean \cap R)},$$

which from the tree is

$$P(Choc\,|R) = \frac{1/5}{1/5 + 0} = 1.$$

Or use the second tree.

(d) The events "Red" and "Chocolate" are independent if and only if

$$P(R \cap Choc) = P(R)P(Choc).$$

But from the tree, $P(R) = 6/30 = 1/5$, while

$$\begin{aligned} P(Choc) &= P(R \cap Choc) + P(B \cap Choc) + P(G \cap Choc) \\ &= \frac{1}{5} + \frac{1}{10} + \frac{1}{3} \\ &= \frac{19}{30}. \end{aligned}$$

So the question is whether

$$\frac{1}{5} = \frac{1}{5} \times \frac{19}{30},$$

which is surely false.

Equivalently, R and C are independent if and only if $P(R|C) = P(R)$. We've already computed these in part (a), and no, they weren't equal.

While it's a little extra work, it may be worthwhile to build a probability table from one of the trees. This is as follows (with the marginal probabilities, the sum of the rows and columns, filled in):

| | R | B | G | |
|---|---|---|---|---|
| Choc | 1/5 | 1/10 | 1/3 | 19/30 |
| Clean | 0 | 1/5 | 1/6 | 11/30 |
| | 1/5 | 3/10 | 1/2 | 1 |

The translation from a tree like Figure 1 to a table is immediate; you're entering the probabilities you entered at the far right of the tree. (If you were wise.)

□

**Problem 2** (20 pts). *One shot by a standard Klingon phaser has a 20% probability of causing 0 hit points, a 30% probability of causing 1 hit point, and a 50% probability of causing 2 hit points.*

(a) *Find the expected value and standard deviation of the number $Y$ of hit points for one shot by a Klingon phaser.*

(b) *The Klingon cruiser* Sklaar's Liver *has the new Borellium phasers, which cause 25% more damage in hit points than standard phasers. (For example, when a standard phaser causes 2 hit points damage, the Borellium phaser causes $2 \times 1.25 = 2.5$ hit points damage.) Find the expected value $\mu$ and standard deviation $\sigma$ of the number of hit points for one shot by* Sklaar's Liver.

(c) *The Starship* Enterprise *can take up to 60 hit points before being destroyed. Find the probability that a Klingon firing standard phasers will destroy the* Enterprise *with 40 independent shots. Is this number exact or approximate? Explain in the space below the answer boxes.*

*Solution.* You're given the following table for the standard Klingon phaser:

| $x$ | 0 | 1 | 2 |
|---|---|---|---|
| $P(X = x)$ | 0.2 | 0.3 | 0.5 |

(a) The expected value of $X$ is $0 \times 0.2 + 1 \times 0.3 + 2 \times 0.5 = 1.3$. The standard deviation can be computed as $E(X^2) - \mu^2$, and

$$E(X^2) = 0^2 \times 0.2 + 1^2 \times 0.3 + 2^2 \times 0.5 = 2.3,$$

thus $\sigma^2 = E(X^2) - \mu^2 = 2.3 - 1.3^2 = 0.61$, or finally

$$\sigma = 0.781$$

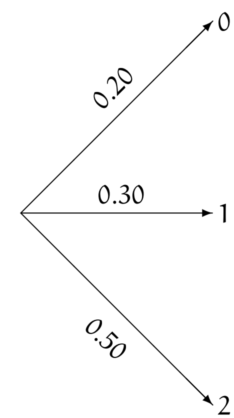
to three decimals (*as required by the instructions*).

(b) The random variable which counts hit points for Borellium phasers is exactly 1.25 times the random variable which counts hit points for standard phasers; therefore its expected value and standard deviation are 1.25 times those quantities for standard phasers:

$$\mu = 1.25 \times 1.3 = 1.625, \qquad \sigma = 1.25 \times 0.7810 = 0.976,$$

rounded to three decimals.

(c) The probability can be computed exactly: all you have to do is build the three-level tree,

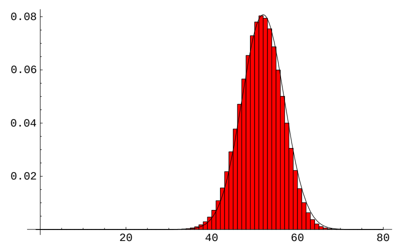



and then build copies off each branch, up to 40 levels, and keep track of hit-points as you travel through the tree. That last level is a doozy: it has $3^{40} = 12157665459056928801$ nodes!

OK, so maybe that isn't practical. It can be shown that the probability of getting exactly $n$ hit-points is the coefficient of $x^n$ in $(0.2 + 0.3x + 0.5x^2)^{40}$. While you can't do this by hand in any reasonable time, computer programs like *Mathematica* can do it in no time (well, 0.069474 second on my MacBook Pro): and the sum of the coefficients of $x^{60}, x^{61}, \ldots, x^{80}$ is

$$0.0616778.$$

This answer is exact to seven decimal places, and is what *you* should aspire to as *your* answer. Also, it leads to the following graph: the red columns have as their height the *exact* probability of getting $x$ hit-points; we've overlaid that with the normal curve with mean 52 and standard deviation 4.93964 (see below for the explanation of why).




But obviously none of this is possible for a student taking the exam. What did we have in mind? The answer is, *We wanted you to apply the Central Limit Theorem.* If $X_1$ denotes the hit points of the first shot, $X_2$ the hit points of the second shot, through $X_{40}$, and we set

$$X = X_1 + X_2 + \cdots + X_{40},$$

then the Central Limit Theorem says that X is *approximately* normally distributed. Since

$$E(X) = 40 \times 1.3 = 52, \qquad \sigma_X = \sqrt{40} \times 0.781025 = 4.93964,$$

(note that we had to use the standard deviation of the $X_i$ to *more than three-digit accuracy* in order to get the standard deviation of X to three-digit accuracy), we can approximate

$$P(X \geq 60) = P\left(\frac{X-52}{4.93964} \geq \frac{60-52}{4.93964}\right) = P(Z \geq 1.61955).$$

Approximating 1.61955 by 1.62, we find that

$$P(0 \leq Z \leq 1.62) = 0.4472,$$

hence

$$P(Z \geq 1.62) = 0.5 - 0.4472 = 0.0528.$$

Rounded to three decimals, we were looking for 0.053 as your answer.

Compared to the actual exact value of 0.0616778, this isn't very impressive. It can be improved quite a bit using the *continuity correction*. Since not all of you were taught the continuity correction, we didn't require you to use it; it consists of computing

$$P(X \geq 59.5) \approx P\left(Z \geq \frac{59.5-52}{4.93964}\right) = P(Z \geq 1.51833).$$

Since $P(Z \geq 1.52) = 0.0643$, this is a much better approximation to the true value.

We didn't ask for the probability that *Sklaar's Liver* could destroy the *Enterprise* in 40 shots, although some students calculated it. (The problem explicitly says, "standard phasers". Still, we gave most of the credit for such an answer.)

But there is an interesting lesson in such a calculation. With standard phasers, the expected damage from 40 shots is $40 \times 1.3 = 52$ hit points—less than the amount it takes to destroy the *Enterprise*. We're not surprised that the probability of destroying the *Enterprise* is only 0.0617. With Borellium phasers, which are only 25% more efficient, the probability of destroying the *Enterprise* turns out to be 0.8191, more than *thirteen times* as great. It might be well to remember this sort of calculation the next time the Pentagon asks for a fighter which commentators remark is "only" 25% better than the current ones. Sometimes that 25% can buy you a 1300% improvement in killing capacity.

Finally, there's a subtlety that we worried about whether students would notice (we were prepared to grade the problem correct either way): exactly what happens in part (c) if the Klingon scores exactly 60 hit-points? Is the *Enterprise* destroyed, or not? The English in "can take up to 60 hit points before being destroyed" is a little vague. Perhaps one should calculate $P(X \geq 61)$ rather than $P(X \geq 60)$, on the grounds that 60 hit points aren't enough. To find the answer to that you'll have to read the training manuals for the *Enterprise*, which unfortunately are classified... (No one noticed the subtlety.)

□

**Problem 3** (20 pts). *On any given day, Jane has in her refrigerator her favorite bottles of fruit juices: orange and/or apple. Let X* *denote the number of bottles of orange juice and* $Y$ *denote the number of bottles of apple juice on any given day. The joint probability distribution of* $X$ *and* $Y$*,* $P_{XY}(x, y)$*, is given by:*

| | | X | | | |
|---|---|---|---|---|---|
| | | *1* | *2* | *3* | |
| | 0 | 0.075 | 0.1 | 0.025 | |
| Y | 1 | 0.15 | 0.15 | 0.30 | |
| | 2 | 0.125 | 0.05 | 0.025 | |
| | | | | | |

(a) *Fill in the table the marginal distribution of* $X$ *and the marginal distribution of* $Y$*.*
(b) *Find the expected value of* $X$*, the expected value of* $Y$*, and the standard deviation of* $X$*.*
(c) *Given that on a certain day she has no apple juice in her refrigerator, find the probability that on that day she has at most two bottles of orange juice in her refrigerator.*
(d) *Find* $\mathrm{Cov}(X, Y)$*.*

*Solution.*

(a) "Marginal distribution" just means: add the rows and columns. Here they are:

| | | X | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | |
| | 0 | 0.075 | 0.1 | 0.025 | 0.20 |
| Y | 1 | 0.15 | 0.15 | 0.30 | 0.60 |
| | 2 | 0.125 | 0.05 | 0.025 | 0.20 |
| | | 0.35 | 0.30 | 0.35 | 1.00 |

Separating them out (which you weren't asked to do, but it makes things clearer) we have computed the distribution for $X$,

| $x$ | 1 | 2 | 3 |
|---|---|---|---|
| $P(X = x)$ | 0.35 | 0.30 | 0.35 |

and for $Y$,

| $y$ | 0 | 1 | 2 |
|---|---|---|---|
| $P(Y = y)$ | 0.20 | 0.60 | 0.20 |

(b) So we can compute the mean for $X$,

$$E(X) = 1 \times 0.35 + 2 \times 0.30 + 3 \times 0.35 = 2.00,$$

the mean of $Y$,

$$E(Y) = 0 \times 0.20 + 1 \times 0.60 + 2 \times 0.20 = 1.00.$$

Finally we compute the variance of $X$,

$$\begin{aligned} \sigma^2 &= E(X^2) - \mu^2 \\ &= 1^2 \times 0.35 + 2^2 \times 0.30 + 3^2 \times 0.35 - 2^2 \\ &= 4.7 - 4 = 0.7, \end{aligned}$$

so that

$$\sigma_X = \sqrt{0.7} = 0.837$$

to three decimal places *(as required)*.

While the textbook seems to prefer horizontal tables to summarize the probability distribution, *vertical* tables are more convenient for the user, mostly because we're used to adding *columns* of figures. Consider the problem of computing $E(X)$ and $\sigma_X$, for example. We fill out the table as follows:

| $x$ | $P(X = x)$ | $x \cdot P(X = x)$ | $x^2 \cdot P(X = x)$ |
|---|---|---|---|
| 1 | 0.35 | 0.35 | 0.35 |
| 2 | 0.30 | 0.60 | 1.20 |
| 3 | 0.35 | 1.05 | 3.15 |
| | | 2.00 | 4.70 |

from which we read off $\mu_X = E(X) = 2.00$ and $E(X^2) = 4.70$, allowing us to compute $\sigma_X^2 = E(X^2) - \mu_X^2$ as above.

Why does the book prefer horizontal tables? No doubt because they use less space when typeset. But what's convenient for the publisher may be very different from what's convenient for the student.

(c) $Y$ was the number of bottles of apple juice, and *given that* she has no apple juice on a given day means that we're on the first row of the table:

| | $X=1$ | $X=2$ | $X=3$ | Sum |
|---|---|---|---|---|
| $Y=0$ | 0.075 | 0.10 | 0.025 | 0.20 |

"At most two" bottles of orange juice means $X \leq 2$. So the answer is the proportion of the sum of the entries for $X=1$ and $X=2$ in row 1, relative to the whole row $Y=0$:

$$P(X \leq 2 | Y=0) = \frac{0.075+0.1}{0.2} = \frac{0.175}{0.2} = 0.875.$$

Looking at it another way,

$$\begin{aligned} P(X \leq 2|Y=0) &= \frac{P\left((X \leq 2)\&(Y=0)\right)}{P(Y=0)} \\ &= \frac{0.075+0.1}{0.075+0.1+0.025} \\ &= 0.875. \end{aligned}$$

But the common-sense interpretation—the first one—is likely to stay with you longer than this one. (Much as it would warm the cockles of our hearts to have you remember how to do the second calculation five years from today, we suspect you won't.)

(d) This part involves the most work of the whole problem. *Usually*, the easiest way to compute covariance is by the formula

$$\mathrm{Cov}(X,Y) = E(XY) - \mu_X\mu_Y,$$

and if we do that here, the expectation $E(XY)$ turns out to be

$$\begin{aligned} &1 \times 1 \times 0.15 + 1 \times 2 \times 0.15 + 1 \times 3 \times 0.3 \\ &\quad + 2 \times 1 \times 0.125 + 2 \times 2 \times 0.05 + 2 \times 3 \times 0.025, \end{aligned}$$

which works out to $E(XY) = 1.95$. (Notice that there's no point in computing the contribution from the first row, since one of the multipliers is always 0.) Therefore

$$\mathrm{Cov}(X,Y) = 1.95 - 1 \times 2 = -0.05.$$

*However*, in this case the formula

$$\mathrm{Cov}(X,Y) = \sum_{x,y}(x-\mu_X)(y-\mu_Y)P\left((X=x)\&(Y=y)\right)$$

(which is the actual *definition* of covariance) is a little less work, mostly because $\mu_X$ and $\mu_Y$ happen to be integers. First we build the tables for $X - \mu_X$ and $Y - \mu_Y$,

| $Y-\mu_Y$ \ $X-\mu_X$ | $-1$ | 0 | 1 |
|---|---|---|---|
| $-1$ | 0.075 | 0.1 | 0.025 |
| 0 | 0.15 | 0.15 | 0.30 |
| 1 | 0.125 | 0.05 | 0.025 |

and now the covariance computes to

$$\begin{aligned} \mathrm{Cov}(X,Y) &= (-1) \times (-1) \times 0.075 + (-1) \times 1 \times 0.025 \\ &\quad + 1 \times (-1) \times 0.125 + 1 \times 1 \times 0.025 \\ &= 0.075 - 0.025 - 0.125 + 0.025 \\ &= -0.05. \end{aligned}$$

There's less arithmetic in this version, but who knew? If the means hadn't turned out to be so nice, there might have been a lot more.

□

**Problem 4** (20 pts). (a) *Among 30 distinct stocks available, exactly four stocks have prices that will increase in the next trading period. An investor selects ten distinct stocks at random without replacement. Find the probability that exactly three of the selected stocks have prices that will increase in the next trading period.*

(b) *One of the 30 stocks is AAPL stock. Each trading period, its price has a 60% chance of increasing. In 10 trading periods, find the probability that AAPL's stock price increases in exactly 6 of the trading periods.*

(c) *(Referring to part (b)) In 100 trading periods, find the expected value and standard deviation of the number $Y$ of trading periods in which AAPL's stock price increases.*

(d) *(Referring to part (b)) In 100 trading periods, find an approximate probability that AAPL's stock price increases in 55 to 65 trading periods, inclusive.*

*Solution.*

(a) This is a hypergeometric problem: the larger population, 30, is divided between those which will increase (4) and those that won't (26); the subpopulation, 10, is divided between those which will increase (3) and those which won't (7). So the answer is

$$p = \frac{\binom{4}{3}\binom{26}{7}}{\binom{30}{10}} = \frac{4 \times 647800}{30045015} = \frac{160}{1827}.$$

Rounded to three decimal places, this is 0.088.

(b) This is a binomial probability:

$$\binom{10}{6} \times 0.6^6 \times 0.4^4 = 210 \times 0.046656 \times 0.0256 = 0.250823.$$

You're supposed to round this to 0.251.

(c) This is the expected value and standard deviation of a binomial probability with $n = 100$ and $p = 0.6$:

$$\mu = np = 60,$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{100 \times 0.6 \times 0.4} = \sqrt{24} = 4.89898.$$

$\sigma$ should be rounded to 4.899.

(d) We are asking for $P(55 \le X \le 65)$, where X is binomially distributed with mean 60 and standard deviation 4.89898. Since X is approximately normally distributed, we estimate this as

$$\begin{aligned} P(55 \le X \le 65) &= P(-5 \le X - 60 \le 5) \\ &= P\left(\frac{-5}{4.89898} \le \frac{X-60}{4.89898} \le \frac{5}{4.89898}\right) \\ &\approx P(-1.02062 \le Z \le 1.02062). \end{aligned}$$

We estimate this as $2 \times P(0 \le Z \le 1.02) = 2 \times 0.3461 = 0.6922$, which rounds to 0.692.

Those who know how to take the continuity correction—it wasn't required—will get a more accurate answer:

$$\begin{aligned} P(54.5 \le X \le 65.5) &= P(-5.5 \le X - 60 \le 5.5) \\ &= P\left(\frac{-5.5}{4.49898} \le \frac{X-60}{4.49898} \le \frac{5.5}{4.49898}\right) \\ &= P(-1.2225 \le Z \le 1.2225), \end{aligned}$$

which, if we approximate by 1.22, yields

$$P(-1.22 \le Z \le 1.22) = 2 \times 0.3888 = 0.7776.$$

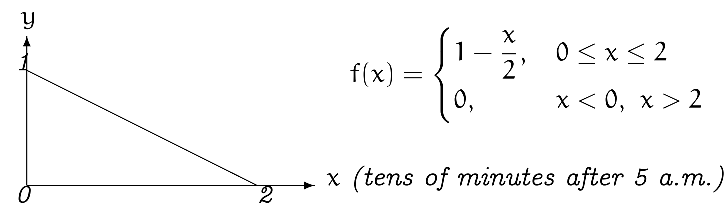
The actual value is

$$\sum_{i=55}^{65} \binom{100}{i} 0.6^i 0.4^{100-i} = 0.738573.$$

Once again, the continuity correction value (0.778) is closer to the true answer (0.739) than the uncorrected value (0.692).

□

**Problem 5** (20 pts). *A particular daily newspaper is delivered between 5 a.m. and 5:20 a.m. The arrival time, X (in tens of minutes after 5 a.m.), of the newspaper has the following probability density function* $f(x)$*:*




$$f(x) = \begin{cases} 1 - \frac{x}{2}, & 0 \le x \le 2 \\ 0, & x < 0,\ x > 2 \end{cases}$$

(a) *Find the probability that on a randomly selected day, the newspaper will arrive after 5:15 a.m.*

(b) *Find the average arrival time of the newspaper.*

(c) *When should a reader look for the newspaper, in order to have a 96% chance that the paper will have arrived?*

*Solution.*

(a) 5 : 15 is 1.5 tens of minutes after 5 : 00. Accordingly, you are being asked for

$$\int_{1.5}^{2} f(x)\,dx = \int_{1.5}^{2} 1 - \frac{x}{2}\,dx = \left(x - \frac{x^2}{4}\right)\Big|_{1.5}^{2} = \frac{1}{16}.$$

You don't have to do this by an integration, of course; the answer is the area of a triangle with base 1/2 and height $f(3/2) = 1/4$, hence is

$$\frac{1}{2}bh = \frac{1}{2}\frac{1}{2}\frac{1}{4} = \frac{1}{16}.$$

Rounded, the answer is 0.062, but this is subtle: the exact answer is 0.0625, but the rule is "round to even" when the last digit is a 5, so you round to 0.062 (**not** 0.063).

(b) The average arrival time of the newspaper is

$$\int_0^2 x f(x)\,dx = \int_0^2 x - \frac{x^2}{2}\,dx = \frac{2}{3},$$

but this is in tens of minutes; so the average arrival time of the newspaper is $5:06.667$, i.e. $5:06:40$.

The physics and engineering majors among you—wait, what are THEY doing among you?!—will recognize that the center of mass of a triangular figure (made of a uniformly dense material) is located at the average of the three vertices. In our case these are at $(0,0)$, $(0,1)$, and $(2,0)$; thus the center of mass is at

$$\frac{(0,0)+(0,1)+(2,0)}{3} = \left(\frac{2}{3}, \frac{1}{3}\right).$$

The $x$ coordinate of this is $\mu_X$.

This rule for finding a center of mass *only works for triangles,* however. It doesn't work for quadrilaterals or higher-order polygons. Those are worth learning for physics and engineering majors, but business majors? You'll never see it again in your life...

(c) We're asking for the value of $c$ so that

$$\int_0^c f(x)\,dx = 0.96,$$

i.e.,

$$\int_0^c 1 - \frac{x}{2}\,dx = 0.96.$$

This is a quadratic equation,

$$c - \frac{c^2}{4} = 0.96,$$

which has two solutions, $c = 1.6$ and $c = 2.4$. Only the first is meaningful (between 0 and 1). Thus the correct answer is 1.6 tens of minutes after $5:00$, i.e. $5:16$.

□

**Problem 6** (20 pts). *A survey was conducted to measure the number of hours per week adults in the United States spend on home computers. In the survey, the number of hours was normally distributed, with a mean of 14 hours and a standard deviation of 2 hours.*

(a) *A survey participant is randomly selected. Find the probability that the number of hours spent on the home computer by the participant is between 11.5 and 19 hours per week.*

(b) *Four survey participants are randomly selected. Find the probability that their average number of hours per week spent on the home computer is less than 13 hours.*

(c) *Five survey participants are randomly selected. What is the probability that among these five surveyed, exactly two spent between 11.5 and 19 hours per week on the home computer?*

*Solution.*

(a) You're being asked for

$$\begin{aligned} P(11.5 \le X \le 19) &= P(-2.5 \le X - 14 \le 5) \\ &= P\left(\frac{-2.5}{2} \le \frac{X-14}{2} \le \frac{5}{2}\right) \\ &= P(-1.25 \le Z \le 2.5) \\ &= P(0 \le Z \le 2.5) + P(0 \le Z \le 1.25) \\ &= 0.4938 + 0.3944 = 0.8882, \end{aligned}$$

which rounds to 0.888.

(b) Let $\overline{X} = (X_1+X_2+X_3+X_4)/4$ be the average of four randomly selected survey participants. Then $E(\overline{X}) = 14$ but $\sigma(\overline{X}) = \sigma/2 = 1$. Thus

$$\begin{aligned} P(\overline{X} < 13) &= P(\overline{X} - 14 < -1) \\ &= P(Z < -1) \\ &= 0.5 - P(0 \le Z \le 1) \\ &= 0.5 - 0.3413 = 0.1587. \end{aligned}$$

That rounds to 0.159.

(c) From part (a) we see that the probability of *one* spending between 11.5 and 19 hours on the computer is 0.8882. Therefore the probability that exactly two of five will do this is

$$\binom{5}{2}0.8882^2 0.1118^3 = 0.011$$

rounded to three decimals. Why did we use *four* decimals in the calculation? Would it have made any difference if we had used three?

□

**Problem 7** (20 pts). *The time required to complete a data entry in a particular computerized accounting system is an exponentially distributed random variable with a mean of 2 minutes.*

(a) *Find the probability that a randomly selected data entry will be completed within 1.6 minutes.*

(b) *Assume that the sequence of times between consecutive data entries is a sequence of independent and identically distributed exponential random variables.*

  (i) *Find the expected value and standard deviation of the number, X, of entries completed within an 8-minute period.*

  (ii) *On an occasion when there are five entries to be input, find the probability that exactly five entries will be completed within an 8-minute period.*

*Solution.*

(a) The PDF of an exponentially distributed random variable is of the form $\lambda e^{-\lambda x}$, where the mean value is $1/\lambda$; therefore for *this* exponentially distributed RV, we have $\lambda = 1/2$ (data entries per minute).

We have

$$P(X \leq t) = 1 - e^{-\lambda t} = 1 - e^{-t/2}.$$

Substituting $t = 1.6$ we get

$$P(X \leq 1.6) = 1 - e^{-0.8} = 0.551.$$

(b) Since the times *between* transactions are exponentially distributed and independent, the *number of transactions* in any particular unit of time is Poisson-distributed.

  (i) You're being asked to specify the unit time interval as 8 minutes. Since on average 2 minutes are required per transaction, this means an average of 4 transactions will be completed in 8 minutes. If X denotes the number of transactions actually completed, therefore

$$P(X = k) = \frac{4^k}{k!}e^{-4}. \tag{1}$$

Now *in general*, if

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

the mean of X is $\lambda$ and the standard deviation of X is $\sqrt{\lambda}$. *In this case*, therefore, the mean of X is 4 (which makes sense: we *said* we averaged 4 transactions per 8-minute period), and the standard deviation is $2 = \sqrt{4}$.

  (ii) From the formula (1), we find

$$P(X = 5) = \frac{4^5}{5!}e^{-4} = \frac{128}{15}e^{-4} \approx 0.156$$

to three decimals. (Either answer, to three decimals or as a rational multiple of $e^{-4}$, was accepted.)

□

**Problem 8** (20 pts). *In an attempt to measure the gestation period of ferrets, thirteen females were observed. The gestation periods for these 13 females were found to be*

40 43 43 42 43 43 43 43 45 45 44 40 38

*days, respectively. In what follows, it may help if I tell you that* $\sum_i X_i = 552$ *and* $\sum_i (X_i - \overline{X})^2 = 49.2308$.

(a) *Find the mean* $\overline{X}$ *and the sample standard deviation* $s$ *of this sample.*

(b) *Find the 95% confidence interval for the mean gestation period of ferrets.*

(c) *Later it is determined that the standard deviation of the gestation period is actually* $\sigma = 1.95$ *days. Find the 95% confidence interval for the mean gestation period using the above data and this value of* $\sigma$.

*Solution.* Actually, the mean gestation period for ferrets really is about 40 days. I know this because I once knew a couple who owned a ferret. (Which is now illegal in the State of California, incidentally.) As far as I'm concerned, I'd just as soon own a feral rat.

In the published version of the exam—the exam actually handed out to students—there was a misprint: there were 14 data items instead of 13. Students were instructed to remove one of the 45's from the list (we've already done that for you in this solution set). However, this was completely irrelevant: all that matters is the sums $\sum_i X_i$ and $\sum_i (X_i - \overline{X})^2$.

(a) The mean is given by
$$\overline{X} = \frac{1}{n}\sum_i X_i = \frac{552}{13} = 42.462,$$
while
$$s^2 = \frac{1}{n-1}\sum_i (X_i - \overline{X})^2 = \frac{49.2308}{12} = 4.10257.$$
Therefore
$$s = \sqrt{4.10257} = 2.025$$
to three decimals.

(b) Because we have a *sample* standard deviation, and used the $n-1$ in the denominator, we must use the t-statistic. The 95% confidence interval is therefore
$$\overline{X} \pm t_{0.025,12}\frac{s}{\sqrt{13}}.$$
Looking up in the t table we find $t_{0.025,12} = 2.179$, hence the confidence interval is
$$42.462 \pm 1.224 \text{ days},$$
i.e. between 41.238 days and 43.686 days.

(c) If we actually know $\sigma$, then the confidence interval is obtained using the normal distribution, and is
$$\overline{X} \pm Z_{0.025}\frac{\sigma}{\sqrt{n}},$$
i.e.
$$42.462 \pm 1.960 \times \frac{1.95}{\sqrt{13}} = 42.462 \pm 1.060 \text{ days}.$$

□

**Problem 9** (20 pts). *Maserghini Auto, Inc. claims that its cars from the latest model year can accelerate from* 0 *to* 60 *miles per hour in* 4 *seconds or less, on average. CarTrend Magazine decides to test this claim by timing* 36 *randomly-chosen Maserghinis from this year. Suppose the average* 0-60 *time for these* 36 *cars times is* 4.1 *seconds, with a sample standard deviation of* 0.24 *seconds.*

(a) *Formulate appropriate null and alternative hypotheses for the acceleration times of the cars.*
(b) *Choose an appropriate test statistic and describe the rejection region for the 2.5% significance level.*
(c) *At the 2.5% significance level, should CarTrend conclude that Maserghini's claim is false?*
(d) *Based on the appropriate attached table, find an interval containing the* $p$*-value.*

*Solution.*

(a) $H_0 : \mu = 4$ (or $\mu \leq 4$).
$H_a : \mu > 4$.

(b) Since we are given a *sample* standard deviation, we must use the t-test. That means we use the t-statistic. The rejection region for the 2.5% significance level is
$$t > t_{0.025,35} = 2.030.$$

(c) We compute the t-statistic by
$$t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}} = \frac{4.1 - 4.0}{0.24/\sqrt{36}} = 2.5.$$
By the rejection rule in part (b) above, since $2.5 > 2.030$, we *do* reject at the 2.5% level.

(d) We reproduce the entire row of the t-cutoff table for 35 degrees of freedom:

| $\alpha$ | 0.1 | 0.05 | 0.025 | 0.01 | 0.005 | 0.001 |
|---|---|---|---|---|---|---|
| $t_{\alpha,35}$ | 1.306 | 1.690 | 2.030 | 2.438 | 2.724 | 3.340 |

Wke're asking which entries are smaller than 2.5; for those which are, we *can* reject at that level; for those which aren't, we *can't*. From the table we see that we *can* reject when $\alpha = 0.01$, but we *can't* when $\alpha = 0.005$. Therefore the p-value lies in the interval $(0.005, 0.01)$. A calculator or computer would allow us to calculate the exact p-value: $p = 0.008628$. All we can say with the given tables is that the p-value is somewhere between 0.005 and 0.010.

□

**Problem 10** (20 pts). *A researcher wishes to test the assertion of a particular student that at least 70% of all college students would prefer not to take a class that meets earlier than 10 a.m. For this purpose, the researcher conducts an opinion poll of 100 randomly selected students.*

(a) *Formulate appropriate null and alternative hypotheses for the researcher to use.*
(b) *Choose an appropriate test statistic and describe the rejection region for the 4% significance level.*
(c) *Let $X$ be the number of students who participate in the above poll and prefer not to take classes earlier than 10 a.m. Find the maximal value of $X$ that would allow the researcher to reject the student's assertion, using the test statistic and rejection region in (b).*

*Solution.*

(a) $H_0 : p = 0.7$ or $H_0 : p \geq 0.7$; and $H_a : p < 0.7$. We use this formulation of $H_a$ because it's clear from the context of the problem (especially part (c)) that we doubt the student's assertion.
(b) This is a binomial proportion problem, so we must use the $Z$ statistic. We compute

$$\sigma = \sqrt{\hat{p}(1-\hat{p})/n} = \sqrt{0.7 \times 0.3/100} = 0.0458258,$$

and therefore

$$Z = \frac{\hat{p} - 0.7}{0.0458258}.$$

The rejection region is

$$Z < -Z_{0.04},$$

but $Z_{0.04}$ is not one of the standard cutoffs we use; we'll have to compute it ourselves. We're looking for a value $c$ so that $P(Z > c) = 0.04$; which means $P(0 \leq Z \leq c) = 0.46$. $c = 1.75$ comes pretty close, because $P(0 \leq Z \leq 1.75) = 0.4599$; so we'll take $Z_{0.04} = 1.75$. (To five decimals, the actual value is $Z_{0.04} = 1.75069$, so this is pretty tight.)

Thus our rejection rule becomes

$$Z < -1.75.$$

(c) So we would compute $Z = \frac{\hat{p}-0.7}{4.58258}$, and check for $Z < -1.75$. This translates to

$$\hat{p} < 0.7 - 1.75 \times 0.0458258 = 0.619805.$$

Since $X$ is the number of students, we have $\hat{p} = X/100$; and therefore our rejection rule translates to

$$X < 61.9805;$$

so if 61 students report that they don't like taking classes before 10 a.m. we'll reject the null hypothesis; but if 62 report that, we can't reject. 61 is the maximal value of $X$ which would allow the researcher to reject the student's assertion.

□