Statistics 2 - Summer 2009 - Practice Final Detailed Solutions

1. The most reliable tactic is to find the probability of one possible ordering, then multiply by the number of possible orderings. Let's consider the ordering AAAAB, where A is the rank of your four of a kind, and B is the odd card.

First card can be anything in the deck. There are 52 cards, any of them will work. Second draw, there are 51 cards left, but you want only one of the three remaining cards that match the rank of the first draw. Similar for draws three and four. The final draw there are 48 cards left in the deck. None of them can match the rank of the first four, so it can be any card remaining.

You have five cards. 4 are of one rank, 1 is of another. You use the binomial coefficient to sort out the number of possible orderings.

$\frac{5!}{4!1!} \times \frac{52}{52} \times \frac{3}{51} \times \frac{2}{50} \times \frac{1}{49} \times \frac{48}{48}$

2. The law of averages says that, as the number of draws increases, the observed proportion will converge to the expected proportion. Since our expected proportion is inside the desired range, a larger number of draws will give a higher probability of being inside that range.

3. Think about what the probability histograms might look like. With 60 rolls, there are 61 possible values (0 through 60). With 600 rolls, there are 601. The law of averages says that the observed proportion will be very close to 1/6 with a large number of rolls, but that's not what this problem is about. It's asking about the probability of getting one specific value. You've got to think that the individual bar for one value will have more area when there's only 61 different values than when there are 601. In case you're curious:

60 rolls, P(exactly 1/6 sixes) $= \frac{60!}{10!50!} \times \frac{1}{6}^{10} \times \frac{5}{6}^{50} \approx 14\%$

600 rolls, P(exactly 1/6 sixes) $= \frac{600!}{100!500!} \times \frac{1}{6}^{100} \times \frac{5}{6}^{500} \approx 4\%$

4. The draws are being made with replacement, so we don't need to worry about the correction factor. The question asks only for the EV and SE, which can be calculated without concern for whether or not the normal approximation is appropriate. We're asked about the sum of draws, so the formulas for EV and SE of sum are used.

First step is to find the mean and SD of the box:

Mean of box: $\frac{-10-5+0+1+3+7+9}{7} = \frac{5}{7}$

There are more than two different numbers in the box, so we have to muscle the SD out the long way:

SD of box: deviances: $\frac{-75}{7}\,\frac{-40}{7}\,\frac{-5}{7}\,\frac{2}{7}\,\frac{16}{7}\,\frac{44}{7}\,\frac{58}{7}$
Square: $\frac{5625}{49}\,\frac{1600}{49}\,\frac{25}{49}\,\frac{4}{49}\,\frac{256}{49}\,\frac{1936}{49}\,\frac{3364}{49}$
Mean: $\frac{1830}{49}$
SD = Root : 6.111

Next you use the formulas for EV and SE of a sum to arrive at the solution:

EV $= 100 \times \frac{5}{7} = 71.43$

$\text{SE} = \sqrt{100} \times 6.11 = 61.11$

5. The box, as given, doesn't give us the information we need. We're counting the negative numbers, so we want a box that will have a 1 if the number pulled is negative, and a 0 otherwise:

New box: [1 1 0 0 0 0 0]

We then proceed in the same way as the previous problem. The EV and SE of the count are calculated in the same way as the EV and SE of the sum... a count is like a sum from a box with only zeroes and ones.

Mean of box: $\frac{2}{7}$

Because there are exactly two numbers in the box, we can use the shortcut formula to get the SD:

SD of box: $(1-0)\sqrt{\frac{2}{7} \times \frac{5}{7}} = 0.452$

$\text{EV} = 400 \times \frac{2}{7} = 114.29$

$\text{SE} = \sqrt{400} \times 0.452 = 9.04$

6. Box: [1 0]
Mean of box: 0.5
SD of box: $(1-0) \times \sqrt{0.5 \times 0.5} = 0.5$

$\text{EV} = 6 \times 0.5 = 3$
$\text{SE} = \sqrt{6} \times 0.5 = 1.225$

$z = \frac{3.5-3}{1.225} = 0.41$
Chance $= \frac{100\% - 31.08\%}{2} = 34.46\%$

7. Mean of box: 0.36
SD of box: $(1-0) \times \sqrt{0.36 \times 0.64} = 0.48$

$\text{EV} = 100 \times 0.36 = 36$
$\text{SE} = \sqrt{100} \times 0.48 = 4.8$

8. Mean of box: 0.04
SD of box: $(1-0) \times \sqrt{0.04 \times 0.96} = 0.196$

$\text{EV} = 4\%$
$\text{SE} = \frac{0.196}{\sqrt{25}} \times 100\% = 3.9\%$

The confidence interval cannot be constructed. Going two SEs below the EV gives a negative proportion, which is not a possible value.

9. We want to consider all three colors at once, so the chi-squared test is appropriate.

Null: The box contains 50% red balls, 45% blue balls, and 5% white balls.

Alt: The box contains some other proportion of the colors.

Expected reds: $25 \times .5 = 12.5$
Expected blues: $25 \times .45 = 11.25$
Expected whites: $25 \times 0.5 = 1.25$

$\chi^2 = \frac{(15-12.5)^2}{12.5} + \frac{(9-11.25)^2}{11.25} + \frac{(1-1.25)^2}{1.25} = 1$
3 - 1 = 2 d.f.

There's no significance level stated, so use the default of 5%.

$50\% < p < 70\%$ is not significant at the 5% level, so we fail to reject the null hypothesis. It looks like the draws could come from the box described.

10. The calculations are correct as long as you have a simple random sample. A simple random sample is like draws at random without replacement from a box that has one ticket for everyone in the population. Here, there were many stages... there was an SRS at the bottom, but the boxes they were drawn from aren't the entire population.

11. It makes sense that the viewers of a program likely hold similar opinions to the host of the program, especially if the opinions are extreme (to either wing, right or left.) Further, because it was a call-in poll, the only people who responded were those who were (1) watching the program in the first place and (2) had a strong enough opinion to call in. The figure is likely pretty badly biased.

12. This is a simple random sample, so you know the formulas for finding the EV and SE. A quick double check of the math shows that the stated figures are correct.

13. We're dealing with a deck of cards, so there are 52 total cards. 13 of them are spades, so we can construct a box with zeroes and ones to represent the situation.

Box has 13 ones and 39 zeroes.
Mean of box: 0.25
SD of box: $(1-0)\sqrt{.25 \times .75} = 0.433$

$\text{EV} = 13 \times .25 = 3.25$
When doing the SE, we need the correction factor. There are only 52 items in the box, and we're drawing 13 of them without replacement. That's more than 10%.

$\text{SE} = \sqrt{\frac{52-13}{52-1}} \times \sqrt{13} \times 0.433 = 1.365$

There's only a few possible values, the integers 0 through 13. The width of half a bin in the probability histogram likely contains a fair amount of area, so we need to use the continuity correction. Since we're asked about 5 or more, we want to make sure the entire bin for 5 is included. Thus, we want to move half a step down to 4.5.

$z = \frac{4.5-3.25}{1.365} = 0.90$

$\text{Chance} = \frac{100\%-63.19\%}{2} = 18.405\%$

14. We're asked only about the number of fours, so the box will have a one and five zeroes if the die is fair. This means the count will go up by one if we roll a 4, and the count will be unchanged if anything else is rolled. Rolling a die is like drawing with replacement.

Null: The data are like 60 draws at random with replacement from a box with 1 one and 5 zeroes.
Alt: The box has a larger proportion of ones.

Mean of box: $\frac{1}{6}$
SD of box: $(1-0) \times \sqrt{\frac{1}{6} \times \frac{5}{6}} = .3727$

EV: $60 \times \frac{1}{6} = 10$
SE: $\sqrt{60} \times .3727 = 2.887$

The possible values the count can take are the integers between 0 and 60. Using the normal approximation to get a p-value is fine, as going 2 SEs above or below the EV is still comfortably inside the range of possible values. There's a large number of possible values, so we're not concerned with the width of half a bin in the probability histogram. The continuity correction is not needed.

$z = \frac{15-10}{2.887} = 1.95$
$p = \frac{100\%-94.88\%}{2} = 2.56\%$

p is significant at the 5% level, so we reject the null hypothesis. It looks like there are too many fours.

15. This time, we want to consider all six outcomes at once, so the chi-squared test is appropriate.
Null: The data are like 60 draws at random with replacement from a box with equal proportions of 1, 2, 3, 4, 5, and 6.
Alt: The box has some other proportions of the values.

$\chi^2 = \frac{(9-10)^2}{10} + \frac{(8-10)^2}{10} + \frac{(9-10)^2}{10} + \frac{(15-10)^2}{10} + \frac{(10-10)^2}{10} + \frac{(9-10)^2}{10} = 3.2$
d.f. $=$ 6-1 $=$ 5

$50\% < p < 70\%$ is not significant at the 5% level, so we fail to reject the null hypothesis. It looks like the die is fair.