1. Three cards are dealt from a standard deck of 52. (A deck of cards has 4 suits (clubs, diamonds, hearts, spades) with 13 cards in each suit – 2, 3, . . . , 10, jack, queen, king, ace.) Find the following probabilities.

(a) [5 points] The probability that the third card dealt is an ace.

Solution. There are 4 aces out of 52 cards – the answer is $\frac{4}{52}$.

(b) [5 points] The probability that the third card dealt is an ace, given that the first two are not aces.

Solution. If the first two cards are not aces, there are 4 aces out of the remaining 50 cards – the answer is $\frac{4}{50}$.

(c) [5 points] The probability that all three cards dealt are of the same suit.

Solution. The probability that the second card is of the same suit as the first card is $\frac{12}{51}$. Given that this has happened, the probability that the third card is of the same suit as the first two is $\frac{11}{50}$. The answer is $(\frac{12}{51})(\frac{11}{50}) = \frac{22}{425} \approx 0.052$. Alternatively, the probability that the first three cards are all clubs is $(\frac{13}{52})(\frac{12}{51})(\frac{11}{50})$, and similarly for each of the three other suits, so the answer is $4 \times (\frac{13}{52})(\frac{12}{51})(\frac{11}{50})$.

(d) [5 points] The probability that all three cards dealt are aces.

Solution. $(\frac{4}{52})(\frac{3}{51})(\frac{2}{50})$ – again we break down into the probability of getting an ace on each draw, given that we got aces on the prior draws.

2. I have a biased coin with probability one-third of coming up heads.

(a) [8 points] I toss this coin nine times. What is the exact probability that it comes up heads exactly two times? Give your answer as a single fraction or decimal.

Solution. From the binomial theorem this is $\binom{9}{2}(\frac{1}{3})^2(\frac{2}{3})^7$. We have $\binom{9}{2} = (9 \times 8)/(2 \times 1) = 36$ and so the answer is

$$\frac{36 \times 2^7}{3^9} = \frac{2^2 \times 3^2 \times 2^7}{3^9} = \frac{2^9}{3^7} = \frac{512}{2187} = 0.2341.$$ 

(b) [10 points] I toss this coin 225 times. Estimate the probability I get between 68 and 72 heads, inclusive. Keep careful track of endpoints in your estimate.

Solution. The expected number of heads when we flip this biased coin 225 times is $\frac{225}{3} = 75$; the SE is $\sqrt{1/3 \times 2/3 \times 225} = 5\sqrt{2} \approx 7.07$. We make the continuity correction here, so we want to find the area between 67.5 and 72.5 converted to standard units; those are $(67.5 - 75)/7.07 \approx -1.06$ and $(72.5 - 75)/7.07 \approx -0.35$. From tables this area is about $(0.7063 - 0.2737)/2$, near 22 percent.

Note. It’s not a coincidence that the answers to (a) and (b) are close to each other – if you use the normal approximation on (a) you end up computing the area between the same bounds.
3. A box contains 4 red marbles and 9 blue ones. Four marbles are drawn at random. Find:
   (a) [4 points] the expected number of red marbles drawn.

   **Solution.**
   \[ 4 \times \frac{4}{4+9} = \frac{16}{13} \approx 1.23. \]

   (b) [4 points] the SE for the percentage of red marbles drawn, if the drawing is done with replacement.

   **Solution.** The standard error for the number of red marbles drawn is \( \sqrt{\frac{4 \times 4}{13} \times \frac{9}{13}} = \frac{6}{13} \approx 0.46 \); the standard error for the percentage is \( 0.46/4 \approx 0.12 \).

   (c) [4 points] the SE for the percentage of red marbles drawn, if the drawing is done without replacement.

   **Solution.** Multiply the answer from (b) by the correction factor \( \sqrt{\frac{13 - 4}{13 - 1}} \) to get 10 percent.

4. One part of a large survey involves taking a sample of size 100 from the 10,000 households in a certain district of a city. The organizers of the survey divide this district into five sections, each containing 2,000 households. Then they obtain the sample in two stages. First, they draw two sections at random from the five sections in the district. Second, they draw 50 households at random from each of these two sections. (All draws are made without replacement.)

   For both parts, **answer yes or no and explain your answer.**

   (a) [5 points] Is this sample of 100 households drawn by a probability method?

   **Solution.** Yes – all the selection is made at random, and the interviewers have no discretion about who to choose.

   (Some people pointed out that the survey organizers have the discretion to set up the sections however they want. However this does not create the potential for bias. Also, this is not stratified sampling. Stratified sampling would select a fixed number of people from each section.)

   (b) [5 points] Is it a simple random sample of 100 households from the 10,000 households in the district?

   **Solution.** No. Not every possible sample is equally likely to be chosen. In particular every sample will only contain people from two sections, not all five.

5. Consider the two boxes given below:

   Box A: \[2 \ 4 \ 5\]
   Box B: \[1 \ 3 \ 4 \ 6\]

   (a) One ticket is drawn at random from each box. Find the chance that the number drawn from box A is less than the number drawn from box B.

   **Solution.** There are 3 \times 4 = 12 possible ways to pick a ticket from each box. Of these, 5 of them have the ticket from box \(A\) less than the ticket from box \(B\) – \((2, 3), (2, 4), (2, 6), (4, 6), (5, 6)\). The answer is 5/12.
(b) [5 points] One hundred tickets are drawn at random from box A. What is the chance that between 30 and 40 of these tickets are labelled 4?

Solution. The expected number of 4s drawn is $100/3 = 33.3$ and the SD is $\sqrt{100/3 \times 2/3} = 4.71$. We want the area under the normal curve between $(29.5 - 33.3)/4.71 = -0.81$ and $(40.5 - 33.3)/4.71 = +1.52$. From tables this is about $(0.58 + 0.87)/2 \approx 0.72$.

(c) [10 points] You draw one hundred tickets from box A, with replacement, and add the results. You do the same with box B. Which of the two sums is more likely to be greater than 400: the sum from box A or the sum from box B?

Solution. The expected value of a ticket from box A is 3.67, and the SD is 1.25. The EV of the sum of 100 tickets is therefore 367, with SE 12.5. So 400 is $(400 - 367)/12.5 = 2.64$ SD above average.

The expected value of a ticket from box B is 3.50, and the SD is 1.80. The EV of the sum of 100 tickets is therefore 350, with SE 18.0. So 400 is $(400 - 350)/18.0 = 2.78$ SD above average.

In standard units, 400 is closer to the average in the case of box A – so a sum over 400 is more likely from box A.

6. A letter is drawn 500 times, at random, from the word STATISTICS. There are five offers.

(1) You win a dollar if the number of S’s among the draws is 10 or more above the expected number.

(2) You win a dollar if the number of A’s among the draws is 10 or more above the expected number.

(3) You win a dollar if the number of vowels among the draws is 10 or more above the expected number. (A and I are vowels.)

(4) You win a dollar if the number of consonants among draws is 10 or more above the expected number. (C, S, and T are consonants.)

(5) You win a dollar if the total number of S’s and T’s among the draws is 10 or more above the expected number.

Answer the following questions.

(a) [10 points] Rank these five offers in order of the chance they give of winning. Some may be equally likely using the approximations we usually make. Give reasons for your ranking.

Solution. If the SE of a random quantity is larger, the probability that it is at least 10 above its expected value is larger. (Note that we’re making the normal approximation here.) So we want to rank the random quantities in (1) - (5) by their SE. The quantities in (1) and (3) both have the same SE, $\sqrt{500 \sqrt{3/10 \times 7/10}}$, since in both cases there are three “good” letters out of 10. The SE in (4) is the same as that in (1) and (3) – there are seven “good” letters. The SE in (2), $\sqrt{500 \sqrt{1/10 \times 9/10}}$, is smaller; the SE in (5), $\sqrt{500 \sqrt{6/10 \times 4/10}}$, is larger. So (2) is least likely; (1), (3), and (4) are all approximately equally likely and more likely than (2); (5) is most likely.

(b) [5 points] Are any two of these five events mutually exclusive? Explain why or why
not. If your answer is yes, name two of the events which are mutually exclusive.

**Solution.** (3) and (4) are mutually exclusive; if the number of vowels is above its expected value, then the number of consonants must be below its expected value.

(c) [5 points] Are any two of these five events independent? Explain why or why not. If your answer is yes, name two of the events which are independent.

**Solution.** None of the events are independent. For example, if (1) occurs, then the number of Ss is substantially above average. This makes it more likely that the number of consonants is substantially above average, which is the event in (4); it also makes it more likely that the total number of S’s and T’s is substantially above average, the event in (5). It makes it less likely that the number of A’s or the number of vowels is substantially above average – there is simply less room for letters which are not S. Similar arguments hold for every pair of events.