Statistical Mechanics Assignment 1 Solutions

1. (a) There are four cases to consider:
   i. None of the nine players is called out. This event occurs with probability $(\text{OBP})^9 = (0.340)^9 = 6.07 \times 10^{-5}$.
   ii. Exactly one player is called out.
   Suppose that the first player is called out and the others all reach base safely. This occurs with probability $(1 - (\text{OBP}))(\text{OBP})^8 = (0.660)(0.340)^8 = 1.18 \times 10^{-4}$. However, any of the nine players could be called out, so the probability that exactly one player is called out is nine times this figure, or 0.00106.
   iii. Exactly two players are called out.
   Suppose that the first two players are called out, and the others all reach base safely. This occurs with probability $(1-(\text{OBP}))^2(\text{OBP})^7 = (0.660)^2(0.340)^7 = 2.29 \times 10^{-4}$. The same argument could be made for any pair of players in the team. The number of different pairs is $\text{C}(9, 2) = 36$, so the probability that exactly two players are called out is $36(2.29 \times 10^{-4}) = 0.00824$.
   iv. The last possibility is that the ninth player is the third one called out, i.e. that the team just manages to bat around. Using the same argument as above, the probability that there are exactly two outs among the first eight players is $\text{C}(8, 2)(1 - (\text{OBP}))^2(\text{OBP})^6$ and the probability that the ninth player is called out is $(1 - (\text{OBP}))$, so the probability of this event is $\text{C}(8, 2)(1 - (\text{OBP}))^3(\text{OBP})^6 = 28(0.660)^3(0.340)^6 = 0.0124$.

   The simpler events for which we calculated probabilities above are mutually exclusive. The probability of the event that a team bats around is therefore just the sum of these probabilities, or

$$P(\text{bat around}) = 6.07 \times 10^{-5} + 0.00106 + 0.00824 + 0.0124 = 0.0218,$$

   or roughly once in every 46 innings, the equivalent of five games.

   (b) The probability that a team does not bat around during a game is $[1-P(\text{bat around})]^9 = (1 - 0.0218)^9 = 0.820$. The probability that they do bat around at least once is therefore $1 - 0.820 = 0.180$, or about once every five and a half games.

   Something to think about: Based on the probability of batting around in an inning, I said that this should occur roughly once every five games. However, we just calculated that the probability that this will occur at least once in a game corresponds to a frequency of about once every five and a half games. Where does this discrepancy in frequency estimates come from?

2. (a) For convenience, let's number the urns. The urn with the small opening will be urn 1, the urn with the medium opening will be urn 2, and the urn with the large opening will be urn 3. The child will pick up the balls one at a time. For the large

balls, she will find by trial-and-error that they only fit in urn 3, so there is only one way each of these can be placed in an urn. The medium balls can only go in urns 2 or 3. She therefore has two choices of where to put the first of these she picks up, and two choices for the second. Similarly, she has three choices of where to put each of the small balls. The total number of ways she can put the balls into urns is therefore $(1)(1)(2)(2)(3)(3) = 36$. Out of these 36 ways, the number of ways she can end up with at least one ball of each size in urn 3 can be found as follows:

Medium balls:	
1^{st} in urn 3 and 2^{nd} in urn 2 or 3:	2 possibilities
2^{nd} in urn 3 and 1^{st} in urn 2:	1 possibility
	3 possibilities
Small balls:	
1^{st} in urn 3 and 2^{nd} in any urn:	3 possibilities
2^{nd} in urn 3 and 1^{st} in urns 1 or 2:	2 possibilities
	5 possibilities
	$3 \times 5 = 15$ ways of putting at least one ball of each size in urn 3

The probability that this will happen is therefore

$$P = \frac{15}{36} = \frac{5}{12} \approx 0.42.$$

(b) The only balls that can be in container 1 are the small ones. We therefore only need to worry about those. There are, as we saw above, nine ways the youngster can put two balls in three urns. Of those, four have no balls in urn 1 (two ways she can put the first ball in urns 2 or 3, and similarly for the second ball). Thus, the probability that we have no balls in urn 1 is

$$P = \frac{4}{9} \approx 0.44.$$

(c) I simulated this process in Excel. The basic simulation consisted of four columns of data representing the selections made by the child when she picked up the first small ball, the second small ball, the first medium ball, and the second medium ball. We don't need to consider the large balls because they will automatically end up in urn 3. For the small balls, the Excel formula representing selection is

```
=INT(3*RAND())+1
```

This generates a random integer selected from the set {1,2,3}. For the medium balls, the Excel formula I used is

```
=INT(2*RAND())+2
```

This represents a random selection between urns 2 or 3, the only urns in which the medium balls fit. In the next column of my spreadsheet, I created a logical test for the presence of at least one ball in each urn:

```
=AND(OR(A2=3,B2=3),OR(C2=3,D2=3))        (1)
```

In the final column, I created a test for the absence of balls from urn 1:

```
=AND(A2<>1,B2<>1)        (2)
```

In both of these last two formulas, the row number is of course adjusted for each sample in my simulation. I can then just count the number of cases and calculate the probabilities:

Number of samples:	=COUNT(A2:A20000)
Number with at least one ball of each size in urn 3:	=COUNTIF(E2:E20000,TRUE)
Probability estimate (a):	=I3/I2
Number with no balls in urn 1:	=COUNTIF(F2:F20000,TRUE)
Probability estimate (b):	=I5/I2

Column E contains the results of formula 1, while column F of the spreadsheet contains the results of test 2. The rest of these calculations should be reasonably self-explanatory.

I ran a simulation with 2032 samples. The two probability estimates obtained in one run were $P(a) = 0.43$ and $P(b) = 0.45$, which is in very reasonable agreement with our theoretical values of 0.42 and 0.44, respectively. In principle, we should be able to improve the agreement by running a larger simulation. I tested this by running a second simulation with 10 160 samples, and obtained the following estimates: $P(a) = 0.42$, $P(b) = 0.44$, i.e. the correct probabilities to the number of digits reported.