IAP 2016: Binomial Distribution and Sports Competitions: Some Homework Problems

1. Consider tossing a fair coin N times, where M of the tosses land as heads. Using Matlab's factorial function, write a Matlab program to plot the binomial distribution as a function of M: Plot the binomial distribution for $N = 20, 40$ and 160. Do you notice any qualitative differences between the three plots? Where is the peak of the distribution? How does this relate to the number of trials?

2. Now let's toss an unfair coin, where the probability of flipping a heads is 0.4. Modify your code and plot this distribution for $N = 20, 40$ and 160. Where is the peak of the distribution? How does this relate to the number of flips?

3. Prove the two results claimed in class, that the mean of the binomial distribution is

$$\mu = Np,$$

and that the variance is
$$\sigma^2 = Np(1 - p).$$

4. Now lets move on to simulate Mostellar's model of the world series. The idea is as follows. Mosteller claims that a baseball game can be modelled by assuming that the better team has a probability p of winning. Thus to simulate a game we simply will draw a random number, which is uniformly distributed between $[0, 1]$, and then check if the number is less than p. If it is, then the better team wins; if not then it loses.

   We can do this 7 times and count the number of times in the series that the better team wins, and this is a simulation of a world series.

   The MATLAB code needed to simulate a single world series is given on the next page. Run the world series program 44 times using $p = 0.65$ (Mosteller's number). Count the number of times the losing team wins 0,1,2,3 games. Compare with the results of his paper.

5. Now do the same thing and run it 4400 times. Does the fraction of games that the losing team won change significantly by changing the number of Series played? Were Mosteller's measurements for the first fifty years of the twentieth century were lucky.

6. Identify some reason that you are angry at Mosteller's Model. Think about how you would change the model to assuage your anger. If you have time to implement the change in your model, then please do!

```
% Matlab code for simulating a world series
function win=worldseries(p,N)

% p is the probability that the best team wins
% N is the number of games in the world series--in reality N=7
win=0; % this is a counter that counts the number of times the best team wins

for i=1:N
if rand(1,1) < p   % check to see if a random number is < p
win=win+1;   % if it is, increase counter by one.
end
end
```

We have written the world series as a MATLAB *function*. You should save this file in your directory as worldseries.m. If you then type worldseries(0.6,7) the computer will play a world series. for example,

```
>> worldseries(0.6,7)

ans =

    2

>>
```

In this case the best team lost–it only won 2 games in the world series!!

Another useful function that will be required for the homework isto count the number of games the losing team wins.

We can do this using the following code:

```
M=44  % number of games
p=0.65

for n=1:M    % play M games!
    team1=worldseries(p,7);  % number of time team 1 wins
    team2=7-team1; % number of time team 2 wins
    lose(n)=min(team1,team2);  % number of times that the losing team wins
end


  distribution=hist(lose,[0 1 2 3]);

  % distribution(1)=number of games losing team wins 0 games
  % distribution(2)= lost 1 game
  % distribution(3)=lost 2 games
  % distribution (4) = lost 3 games
```

This MATLAB code uses a number of ideas. In the first loop we play M games– in each game the variable *team1* is the number of times that team1 wins while *team2* is the number of times team2 wins. The number of game that the losing team wins is the minimum of the two, which we are saving in the array *lose*.

The command with the word "hist" then counts the distribution of the number of times the losing team wins. No need to understand how–this is just what hist does. You can check that it got it right by typing "lose" and showing the results of your series and then checking that dist gives the right numbers.

If you type *hist(lose,[0,1,2,3])* MATLAB will plot a histogram of the number of times the losing team won.