CS412 – Machine Learning
Probability, Statistics and MLE
100pts

Samuel Lee
24774

Assume we have a coin for which the probability of getting Heads is P_H while the probability of getting Tails is $P_T = 1-P_H$. For unbiased coins, $P_H = P_T = 0.5$.

The Binomial distribution is the gives us the probability of getting k successes (say Heads event) when we run N independent Bernouilli trials (e.g. a coin toss) each with the probability p_H. I.e. we throw the coin N times and each time the probability of heads is P_H and we count how many times we get Heads (=number of success). Here success could have been Tails as well, with numbers just switched...

1) **20pts –** Using some online Binomial calculators (eg. https://stattrek.com/online-calculator/binomial.aspx), compute the following probabilities (write them to 7 digits of precision).

   P (N=100; p=0.5; k=50) = 0.07958923

   P (N=100; p=0.5; k=40) = 0.01084386

   P (N=100; p=0.5; k=60) = 0.01084386

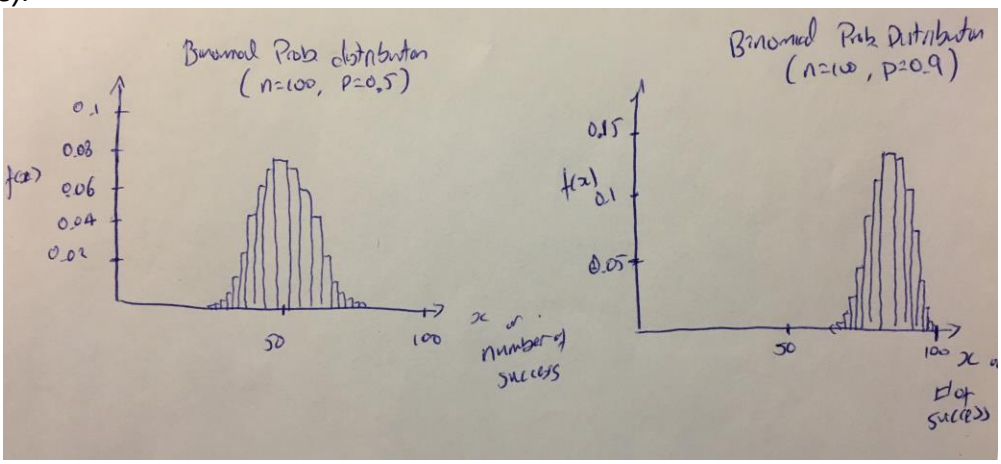
   P (N=100; p=0.9; k=90) = 0.13186534

   P (N=100; p=0.9; k=80) = 0.00117098

   P (N=100; p=0.9; k=100) =0.00002656

2) **30pts –** Roughly plot the two Binomial distributions, one for N=100 and p_H=0.5 and N=100 and p_H=0.9. Note that this is a discrete probability distribution.

Could be by hand, but should be infomative enough (i.e. details such as axis names and a few points should be there).



3) 10pts – If you did not know whether the coin was biased (so you don't know p_H) and you were told that out of the N=100 trials, 90 were Heads, what would be the *best guess* for P_H? The "best guess" is the one as explained in question 5), but an intuitive answer is all that is needed here.

Just one number expected, no explanations needed.

Answer: 0.9

4) 10pts – If you did not know whether the coin was biased (so you don't know p_H) and you were told that out of the N=100 trials, 90 were Heads, how can you find the P_H parameter that maximizes the probability of the data (90 Heads, 10 Tails)? *This method finds the same answer as the intuitive answer in question 3) and has the acronym **MLE**.*

*Just a give the full name for the **formal method of estimation**.*

Answer: Maximum Likelihood Estimation

5) 30pts - We know the formula for the Binomial distribution: $P(N; p; k) = (N \text{ choose } k)p^k.(1-p)^{(N-k)}$

Given the observed data of 90 Heads in 100 trials (so N and k are known), find p that maximizes the probability of seeing this observation (likelihood), using the Binomial distribution formula. You should take the derivative of the likelihood with respect to p and set to 0; and simplify, and then solve for p. *Here p is the p_H mentioned above, simplified and generalized.*

Hint:

- Maximizing the log of the likelihood will be easier.
- You can use: derivative of log x = 1/x

Simplify by removing irrelevant terms and distribute the log:

$$d/dp_H \log (N \text{ choose } k)p^k.(1-p)^{(N-k)} = d/dp_H \log(N \text{ choose } k) + k \log(p) + (N-k) \log (1-p)$$
$$= k/p – (n-k/1-p)$$

For the maximum, the derivative must be 0:

$$k/p – (n-k/1-p) = 0$$

Solve for p = k/n where N=90 and k=100