Assignment 1

Dmitry Adamskiy

October 28, 2018

This assignment is focused on the basics of probability theory and Bayesian inference.

1. Consider a distribution that admits a factorisation of a form $P(a, b, c, d) = P(a)P(b)P(c|a, b)P(d|c, b)$ where all of the variables are binary. How many parameters do you need to specify this distribution? [5 marks]

2. **Three Cards** One card is white on both faces; one is black on both faces; and one is white on one side and black on the other. The three cards are shuffled and their orientations randomized. One card is drawn and placed on the table. The upper face is black. What is the colour of its lower face? (Solve the inference problem.)

   (this exercise is from David J.C. MacKay's book) [10 marks]

3. Solve the following exercises from the textbook:

   (a) **Battleships-1** Exercise 1.20 [10 marks]

   (b) **Battleships-2** Exercise 1.21 [10 marks]

   (c) **Earthquakes-1** Exercise 1.22 [5 marks for each subquestion, 20 in total]

   (d) **Earthquakes-2** Exercise 1.23 [5 marks for each subquestion, 15 in total]

4. **Dunwich Hamlet** A young and purely fictitious football team Dunwich Hamlet [1] has two types of fans. One type is a hardcore football aficionado, who buys season tickets and rarely misses games. The other type is a typical resident of Dunwich village, who supports his local team but does not go to the games very often himself. Dunwich is a family-friendly club and each ticket holder is allowed to bring a family member for free.

   Now the first season comes to an end and the club chairman would like to thank season ticket holders and invite them to a party. But alas the club accounting is not very mature: the database on season ticket sales was lost in an unfortunate accident[2] and no one has ever counted tickets sold on the day.

[1] Not to be confused with Dulwich Hamlet or Dunwich Dynamo.
[2] Dog ate chairman's notebook

However, the chairman holds every copy of Non League Paper which covers Dunwich games and the paper accurately reports the number of spectators for each game. Can he get a good estimate of how many season ticket holders there are from these data?

The model and questions. There are a season ticket holders and b "normal" fans. For each game each fan independently decides whether to go to this game or not. The probability for the season ticket holder to attend the game is p_a and for the "normal" fan it is p_b. The prior distributions on a and b are uniform distributions on $[a_{min}, a_{max}]$ and $[b_{min}, b_{max}]$ respectively. Let c_n be the total number of ticket holders for the n-th game.

(a) Describe the distribution of the number of ticket holders attending the n-th game $p(c_n|a, b)$. What are its mean and variance? [2 marks]

(b) Each ticket holder (independently) with probability p_d brings a family member. Let d_n be the total number of fans at n-th game (this is the number reported in the paper). What is the mean of the distribution $p(d_n|c_n)$? [3 marks]

(c) Draw the graphical model corresponding to this setup (the variables are $a, b, c_1, c_2, \ldots, c_N, d_1, d_2, \ldots, d_N$). [5 marks]

(d) Write a program that computes posterior distributions $P(a|d_1, d_2, \ldots, d_n)$ and $P(b|d_1, d_2, \ldots, d_n)$ and compute these posteriors for $n = 1, 2, \ldots, 10$. Plot these posteriors. What are the ML/MAP estimates of a and b after taking all ten games into account? Use the following model parameters: $a_{min} = 0, a_{max} = 15, b_{min} = 0, b_{max} = 200, p_a = 0.99, p_b = 0.3, p_d = 0.5$. The data sequence $D = [d_1, \ldots, d_{10}]$ is $[22, 27, 26, 32, 31, 25, 35, 26, 28, 23]$. [20 marks]