

5 Data Science (djw1005)

The  $m$  members of the Cambridge University Quiddlefinks Society have been playing matches against each other all term. It's your job as secretary to use match data to estimate the skill level  $s_a$  of each player  $a \in \{1, \dots, m\}$ .

A quiddlefinks match is between two players, and has a single winner. For match  $i \in \{1, \dots, n\}$  we record the two players  $a_i, b_i \in \{1, \dots, m\}$ , and also the winner  $y_i \in \{0, 1\}$ , where  $y_i = 1$  means that player  $a_i$  won and  $y_i = 0$  means that player  $b_i$  won.

- (a) Consider a match between players  $a, b \in \{1, \dots, m\}$ . Suggest a probability model for  $Y = 1_{a \text{ wins}}$  given skill levels  $s_a$  and  $s_b$ . Write down the log likelihood of the full list of match outcomes  $y_1, \dots, y_n$ . [5 marks]
- (b) Explain how to estimate the parameters  $s_1, \dots, s_m$ . Give pseudocode. [3 marks]
- (c) Alternatively we can take a Bayesian approach and let the skill levels  $S_a$ ,  $a \in \{1, \dots, m\}$ , be random variables. Take as prior that skills are independent  $N(0, 1)$ . Explain how to plot the posterior distribution of the skill  $S_a$  of a given player  $a$ . [5 marks]
- (d) You are asked to predict the outcome of an upcoming match between players  $a^*$  and  $b^*$ . Give an expression for the probability that  $a^*$  will win, in terms of the skill levels  $S_{a^*}$  and  $S_{b^*}$ , according to your model from Part (a). [2 marks]
- (e) Explain how to find a 95% posterior confidence interval for the probability you found in Part (d). [5 marks]