ST440/540 – Exam 1 - Due February 20

THIS IS AN EXAM - DO NOT DISCUSS THE PROBLEM WITH ANYONE (INCLUDING OTHER STUDENTS OR THE TA)! If you have questions, please email me.

Cookie Cats is a mobile puzzle app. We analyze the Cookie Cats data available at `https://www.kaggle.com/datasets/yufengsui/mobile-games-ab-testing`. The data were collected by the developers to improves its appeal. The app was deployed to users with the level of the first gate set to either 30 or 40, and the response was one-day retention of the user. Level 30 was given to $n_1 = 44700$ users and $Y_1 = 20034$ were retained; level 40 was given to $n_2 = 45489$ users and $Y_2 = 20119$ were retained.

1. **A/B testing**: Let θ_1 and θ_2 be the true retention probabilities under the two levels. A/B testing is often used to find the optimal setting for an app or website. The final output used for decision making is the posterior probability that each setting is optimal. For cookie cats, we say level 30 is optimal if $\theta_1 > \theta_2$ and level 40 is optimal if $\theta_2 > \theta_1$.

   (a) Describe your approach to computing the posterior distributions of θ_1 and θ_2 in a short paragraph including enough detail that another student could reproduce your work.

   (b) Give the posterior distributions of θ_1 and θ_2, plot them in a single figure and comment differences between them.

   (c) Compute the posterior probability that each level is optimal.

   (d) Conduct a prior sensitivity analysis in a figure or table.

   In addition to the required tables/figures, for parts (b)-(d) include 2-3 sentences describing your methods and summarizing the results.

2. **Bayesian bandit**: Rather than a static experiment as above, we could collect and analyze data sequentially. Say each day we sample 100 users. After the app has been making recommendations and recording data for t days, denote the number of days given level j ($j = 1$ for level 30 and $j = 2$ for level 40) as n_j and the number of successes/retentions under level j as Y_j. Here is some hypothetical data:

Day (t)	1	2	3	4
Level	30	40	40	30
Successes	30	50	40	20
n_1	100	100	100	200
Y_1	30	30	30	50
n_2	0	100	200	200
Y_2	0	50	90	90

You will use Thompson sampling to decide which treatment to give (to all 100 users) each day. On day t, you will draw one sample of θ_1 and one sample of θ_2, denoted θ_1^* and θ_2^*, from their posteriors based on all data collected prior to day t, and if $\theta_1^* > \theta_2^*$ you will give level 30, otherwise you will give level 40. Of course, we cannot get real data, so instead take a subsample (with replacement) of the original data, i.e., if you decide to give level 30 then the data for that day are simulated as $Y \sim \text{Binomial}(1, 100, 20034/44700)$.

(a) Write commented code to implement this sampling scheme. Provide a function that takes the prior hyperparameters and current data as inputs and gives the random treatment assignment as output. The code should be no more than 10 lines. Submit only the function with comments on each line.

(b) Use you code from (a) to carry out this algorithm for 1000 days. Plot the cumulative proportion of days given each level and the posterior means of θ_1 and θ_2, all four plots as a function of t.

(c) Describe the pros and cons of Thompson sampling versus single experiement analyzed in question 1.

Your paper should be written as a professional document with full sentences, clearly labeled figures and tables and few spelling/grammar errors. Organize your report with subsections corresponding to the questions above, i.e., 1a, 1b, ..., 2c. Summarize your analysis in a PDF document that is **no more than two pages long** (12 font, single space, standard margins). Append your code to the end of this document and submit a single document. **In-class students should turn in the exam in class on Monday, Feb 20. Online students should submit the exam on moodle.**

HAVE FUN!