Problemset (3),
Foundations of Machine learning,
HT 2022

Maximilian Kasy

In this problemset you are asked to implement some simulations and estimators in R. Please make sure that your solutions have satisfy the following conditions:

- The code has to run from start to end on the grader's machine, producing all the output.
- Output and discussion of findings have to be integrated in a report generated in R-Markdown.
- Figures and tables have to be clearly labeled and interpretable.
- The findings need to be discussed in the context of the theoretical results that we derived in class.

1. In this problem, you are asked to simulate data for a Bernoulli bandit problem, where

$$D_t \in \{1, \ldots, k\}, \qquad Y_t = Y^{D_t}, \qquad Y_t^d \sim Ber(\theta^d).$$

and treatment is assigned using Thompson sampling with a uniform prior, $(\theta^1, \ldots, \theta^k) \sim U([0,1]^k)$.

   (a) Set up a function which accepts a sample size T and a k-vector $(\theta^1, \ldots, \theta^k)$ as its arguments, and returns a history $(D_t, Y_t)_{t=1}^T$ generated based on the Bernoulli bandit model and Thompson sampling.

(b) Write a second function which takes the same arguments, plus a number of replications R, and evaluates the first function R times (using parallel computing; for instance the *future* package).

This function should return 4 vectors of length T: The averages of Y_t, θ^{D_t}, $\mathbf{1}(D_t = \max \theta^d)$, and $\max \theta^d - \theta^{D_t}$, for each time periord t.

(c) Pick a fixed vector of parameters $(\theta^1, \ldots, \theta^k)$ and a time horizon T and use the second function to plot cumulative average regret as a function of t, using a large number of replications R (such as $R = 10.000$). Repeat this for several different choices of $(\theta^1, \ldots, \theta^k)$.

How does the result relate to the theoretical regret rate bound discussed in class, and to Agrawal and Goyal (2012)?

(d) Now let $k = 2$, fix $\theta^1 = .5$ and $T = 200$. Plot cumulative average regret for T as a function of θ^2, for $\theta^2 \in [0, 1]$. Do the same for the share of observations assigned to the optimal treatment.

How does the result relate to the local-to-zero asymptotics discussed in class, and to Figure 3 in Wager and Xu (2021)?

2. In this problem, we will again consider the Bernoulli bandit, and compare Thompson sampling to exploration sampling, as discussed in Kasy and Sautmann (2021).

(a) Create a modified version of the first function from problem 1, where instead of Thompson sampling treatment is assigned using exploration sampling.

Let this function additionally return the treatment d_T^* with the highest posterior mean.

(b) Create a modified version of the second function from problem 1, again replacing Thompson sampling by exploration sampling.

Let this function additionally return the average policy regret, and the probability of choosing the best arm. Edit the second function from problem 1 to do the same for Thompson sampling.

(c) Pick a fixed vector of parameters $(\theta^1, \ldots, \theta^k)$ and a time horizon T and calculate cumulative average regret as well as average policy regret, for both Thompson sampling and exploration sampling. Do so using a large number of replications R (such as $R = 10.000$).

How does the result line up with the theoretical characterization and simulations of Kasy and Sautmann (2021)?

(d) Repeat this exercise for several different parameter vectors $(\theta^1, \ldots, \theta^k)$ and sample sizes T. Discuss any patterns you might find.

3. In this problem, we repeat the exercise of problem 1, but replace the discrete treatment D_t by a continuous treatment X_t, and replace the uniform prior for θ with a Gaussian process prior for the response function $m(\cdot)$:

$$X_t \in [0,1], \qquad Y_t = m(X_t) + \epsilon_t, \qquad \epsilon_t | X_t \sim N(0,1).$$

Assume that Thompson sampling uses a Gaussian process prior of the form

$$m(\cdot) \sim GP(0, C), \qquad C(x, x') = \tau^2 \exp\left(-\tfrac{(x_1 - x_2)^2}{2\lambda}\right).$$

for $\tau^2 = 4$ and $\lambda = 1/4$.

(a) Write a function which takes a history of observations X_t, Y_t as its argument, and returns a draw from the posterior for $m(\cdot)$, evaluated at a set of grid points $(0, .01, .02, \ldots, 1)$.

Let the function also retorn the maximizer of this posterior draw (over the set of grid points).

(b) Write a second function which accepts a sample size T and a mean function $m(\cdot)$ as its arguments[1], and returns a history $(X_t, Y_t)_{t=1}^T$ generated based on the normal sampling model and Thompson sampling.

(c) Repeat the task of items (b) and (c) from problem 1 for this version of Thompson sampling, for different functions $m(\cdot)$.

You might for instance try polynomials, or trigonometric functions.

(d) Vary the hyper-parameters λ and τ^2, to see how the performance of Thompson sampling is affected.

[1] Recall that you can pass functions as variables in R.

References

Agrawal, S. and Goyal, N. (2012). Analysis of thompson sampling for the multi-armed bandit problem. In *Conference on Learning Theory*, pages 39–1.

Kasy, M. and Sautmann, A. (2021). Adaptive treatment assignment in experiments for policy choice. *Econometrica*, 89(1):113–132.

Wager, S. and Xu, K. (2021). Diffusion asymptotics for sequential experiments. *arXiv preprint arXiv:2101.09855*.