CS6046: Multi-armed bandits
Homework - 3
Course Instructor : Prashanth L.A.
Due : Mar-23, 2018

Theory exercises

1. Let θ denote a univariate parameter and $X_1, \ldots, X_n$ denote i.i.d. samples with Gaussian likelihood, i.e., $p(X_i \mid \theta) = \frac{1}{\sqrt{2\pi}\theta} e^{-\frac{X_i^2}{2\theta^2}}$, for $i = 1, \ldots, n$.
   Answer the following: (1+2+2 marks)
   (a) Work out the posterior update and highlight the form of the posterior density (ignoring the normalization constant).
   (b) Under what choice for the prior is *conjugacy* guaranteed?
   (c) Derive the expression for posterior mean and variance and discuss the asymptotics (i.e., when the number of samples n become large).

2. Consider a two-armed bandit problem. Recall that the ETC algorithm chooses each arm m number of times and then plays the arm with the highest sample mean $(n - 2m)$ number of times. For any horizon n and exploration parameter m (chosen non-adaptively, i.e., before sampling any arm), there exists a problem instance with underlying arms' distribution $v = \mathbb{N}(\mu_1, 1) \times \mathbb{N}(\mu_2, 1)$, such that the regret $R_n(v)$ of ETC on v satisfies
$$R_n(v) \geq cn^{2/3},$$
   where c is a problem-independent constant. (6 marks)

3. Consider a two-armed bandit problem with underlying joint distribution $\nu = p_1 \times p_2$, where p_1 and p_2 are Bernoulli distributions with parameters θ and $1 - \theta$, respectively, for some $\theta \in (\frac{1}{2}, 1)$. Let $v' = p_2 \times p_1$ denote the underlying distribution for a permuted bandit problem. Then, for any bandit algorithm $\mathcal{A}$,
$$\max(R_n(v), R_n(v') \geq \frac{c}{2\theta - 1},$$
   where $R_n(v)$ (resp. $R_n(v')$) is the expected regret with horizon n on problem v (resp. v') and c is a problem-independent constant. (5 marks)

4. Consider a two-armed Bernoulli bandit problem. Suppose that the underlying means are in the set $\{\theta, 1 - \theta\}$ and the bandit algorithm is aware of θ. Does there exist an algorithm $\mathcal{A}$ that satisfies
$$R_n(\mathcal{A}) \leq \frac{c}{2\theta - 1},$$
   where $R_n(\mathcal{A})$ is the expected regret with horizon n and c is a problem-independent constant. If yes, describe the algorithm and derive the regret bound. (7 marks)
   Hint: Try the algorithm in Q5(c) of HW2 or the following variant that uses upper confidence bounds: If the UCB of an arm is better than the optimal mean, play that arm, else alternate between the arms.

Simulation exercise

Consider a ten-armed bandit problem, where each arm's distribution is Bernoulli. Consider the following two problem variants, with respective Bernoulli distribution parameters specified for each arm:

Arms →	1	2	3	4	5	6	7	8	9	10
P1	0.5	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
P2	0.5	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48
P3	0.5	0.2	0.1	No other arms						

Write a program (in your favorite language) to simulate each of the above bandit problems and implement the following bandit algorithms:

- Thompson sampling (TS) with a Beta$(1, 1)$ prior.
- A variant of TS where the prior has mean 0.2 instead of 0.5.
- The UCB algorithm.

Do the following for each problem instance: (12 marks)

1. Choose the horizon n as 10000.
2. For each algorithm, repeat the experiment 100 times.
3. Store the regret in each round $m = 1, \ldots, n$.
4. For TS and its variant, store the (posterior) probability of playing each arm.
5. Plot regret against the rounds $t = 1, \ldots, n$. For TS variants, plot the arm playing probabilities as well.
6. For each plot, add standard error bars.
7. In the figures that report regret performance, plot the gap-dependent lower bound as well as worst case lower bound.

Interpret the numerical results and submit your conclusions. In particular, discuss the following: (3+2 marks)

1. Comparison of the regret performance of TS with Beta$(1, 1)$ prior against that of UCB. How do both algorithm fare when compared to the lower bounds (esp. the gap-dependent one).
2. For the TS variant with a prior mean 0.2, discuss the results, while including comparison to TS with Beta$(1, 1)$ prior.

Here is what you have to submit:

Theory exercises (Q1-4): Hand-written (or typed) answer with concrete justification.

Simulation exercise: Include the following:

- Source code, preferably one that is readable with some comments;
- Plots/tabulated results in a document (or you could submit printouts of plots); and
- Discussion of the results - either hand-written or typed-up.