Bandits with Query Cost

David Krueger

August 2016

We derive conditions under which it is optimal to not query for an N-armed Bernoulli bandit with uniform priors, in terms of the query cost, c, and the discount factor, γ, assuming we are only allowed a single query. Next, we'd like to extend this to non-uniform priors, and cases where multiple queries are allowed.

If we do not query, we achieve expected returns of $\frac{1}{2}\frac{1}{1-\gamma}$. If we query, we (expect to) see 1 or 0 with equal probability $\frac{1}{2}$. Our expected returns after the query are then:

$$\frac{1}{2}(0 + \frac{1}{2}\frac{\gamma}{1-\gamma}) + \frac{1}{2}(1 + \frac{2}{3}\frac{\gamma}{1-\gamma}) - c \tag{1}$$

So now, we just compare these two expectations, and query iff:

$$\frac{1}{2}(0 + \frac{1}{2}\frac{\gamma}{1-\gamma}) + \frac{1}{2}(1 + \frac{2}{3}\frac{\gamma}{1-\gamma}) - c > \frac{1}{2}\frac{1}{1-\gamma} \tag{2}$$

$$\frac{11}{12}\frac{\gamma}{1-\gamma} - \frac{1}{2}\frac{1}{1-\gamma} > c - \frac{1}{2} \tag{3}$$

$$(\frac{11}{12} + c - \frac{1}{2})\gamma > c \tag{4}$$

$$\gamma > \frac{c}{c + \frac{5}{12}} \tag{5}$$

For the case where we only have 1 query, we would only want to query if we knew that it had some chance to change our mind. This means that there must exist some arm, wlog a_1, such that a single observation of a_1 could change it to or from being the highest expectation arm.

When more than one query is allowed, there are more strategies that involve querying, and hence querying might be a good move, even if the cost is larger.