The Thresholding Bandit Problem

Tim Radtke

4/23/2017

The *Thresholding Bandit Problem* described in Locatelli et al. (2016) can be considered a variant of a variant. It can be framed into the wider literature of *Pure Exploration* multi-armed bandit problems. In particular, it shares characteristics with the *Top-m* problem. This problem is concerned with finding the m best arms as described by the means of their corresponding distributions. This in turn is similar to the *combinatorial bandit* problem, which also aims at finding the m best arms. However, it is able to pull several arms at once: Think of an online shop that shows five recommended products on a product detail page. These five recommended products might each be represented by an arm, and we look for the products with the largest mean conversion rate. The thresholding bandit problem we discuss here, however, is concerned with pulling a single arm at a time. And so a more appropriate situation is that of a website presenting a banner. Again, think of an online shop trying to promote a certain category. The content team came up with a number of different designs for the banner, and it's not clear how many click-throughs they will gather.

The idea now is that we would like to classify the banners into two distinct groups: A group with a mean conversion rate μ_i below threshold τ, and a group with mean conversion rate μ_i above threshold τ. This might be the optimization problem when we are concerned with not falling below a certain minimum level click-through rate with the banners we're choosing.

If it turns out that in general we can find relatively quickly whether arms are above or below a threshold, then this kind of test can be used to safeguard against very bad versions in a test. Before we move to a Top-m test, we might want to run a thresholding version and then continue only with the arms classified into the group above the threshold. This might be justified when adjusting parts of the checkout process of an online shop, where the conversion rate or the average order value should not drop below a threshold.

0.1 Setup

In any case, the problem boils down to the following. As standard in multi-armed bandit problems, we have K arms $\mathcal{A} = \{1, \dots K\}$. We can pull arm k at time t to collect feedback in form of the random variable $X_{k,t}$ which is distributed accoring to the arm's distribution ν_k. In general, we are concerned with estimating for each arm i the mean μ_i of its distribution ν_i. In contrast to most bandit problems, however, we do not directly compare arms with each other by comparing their means. In the case of pure exploration bandits, it is for example necessary to compare the arms because one aims to find the best one. In the case of the thresholding bandit, however, we compare each arm's mean μ_i individually against the threshold $\tau \in \mathbb{R}$ which is known and fixed before the experiment. One might compare this to

a pure exploration bandit in which the mean of the best arm is known upfront. We will see later that this fact leads to advantages in the design of algorithms which are not as readily available in other bandit settings.