Learning from interactions

The applicative context relates to studying clinical trials for medical treatments. Let us assume there exist a set of 10 different medical treatments. A new rare disease is discovered, against which we want to evaluate the effectiveness of these 10 treatments. These treatments had been previously developed for other diseases.

Every morning, we can select one (and one only) treatment and apply it to a patient which has been diagnosed with the disease. At the end of the day, we make a test on the patient, which indicates whether the treatment has been effective or not (this is a binary response). The same treatment may sometimes show good results, sometime show ineffective results (still binary), however some treatments may be more likely than others to work. You can model this in a probabilistic way, which differs for each treatment.

Despite this uncertainty, we want to design a strategy (maths/algorithm) to maximise the number of patients cured over 100 days, starting with zero knowledge about the 100 treatments. What happens after the 100 days is not useful.

What you learn from a trial on a patient is assumed to apply to other patients identically. Results from successive trials are assumed to be independent (in the probabilistic sense).

The goal of this short project is to explore and experiment with one of the many topics in the field of machine learning : how to learn from interactions with the environment, in the case the environment provides feedback in the form of rewards (or lack of reward).

1) Design a technique with unsophisticated maths to try to give a good result to the above goal. Describe the technique in 15 to 20 lines. You should identify its good sides and limitations and explain why we can say that this problem needs combination of exploration and exploitation.

2) When you have built your proposal, you should review the following questions :

- what is different in the information conveyed by the two following claims « getting 2 successes of out 4 trials » and « getting 20 successes of out 40 trials » ?
- can we make the difference between the two above claims if we only represent by a proportion (or probability) p_i the knowledge about treatment i ?
- if we now try to represent the information about p_i with a probability distribution over p_i , rather than simply a number, what should be the shape of this distribution in the case we don't know anything about p_i ? What should be its shape in the case we collected a clear majority of successes p_i ?

3) Implement your proposal in Python. You also need to simulate the generator of reply by patients, given the treatment proposed, with a parametric random model. Set your own 10 numerical values for the probabilities that a given treatment will succeed, but of course your trial procedure should not be aware of these values. Display experimental results showing graphically the increase of the number of successes over the 100 days. Your implementation should make a statistical evaluation over multiple 100 days runs to compensate for the variability of the random values of the reply generator. Your software should also be designed to that you will be able to add another strategy to your and compare them.