Random and Two – Armed Bandit Strategy, Patient Treatment

Manav Prabhakar

December 16, 2020

***This document contains both the mathematical part as well as the simulation results.**

Problem Analysis:

Probabilities for two treatments to be successful have been given. The patients to whom the treatments are to be given are independent. We need to analyze two different strategies – random strategy and two – armed bandit strategy. Markov chain models need to be drawn for the latter and the results need to be simulated for verification.

Solution

Consider the following events: -

E_A: Treatment A was chosen; E_B: Treatment B was chosen

E_1: Treatment A is successful; E_2: Treatment B is successful

S: Overall Success

Then, we have: -

$$P(E_A) = 0.5$$

$$P(E_B) = 0.5$$

$$P(E_1/E_A) = \alpha$$

$$P(E_2/E_B) = \beta$$

Using Theorem of total probability, we have

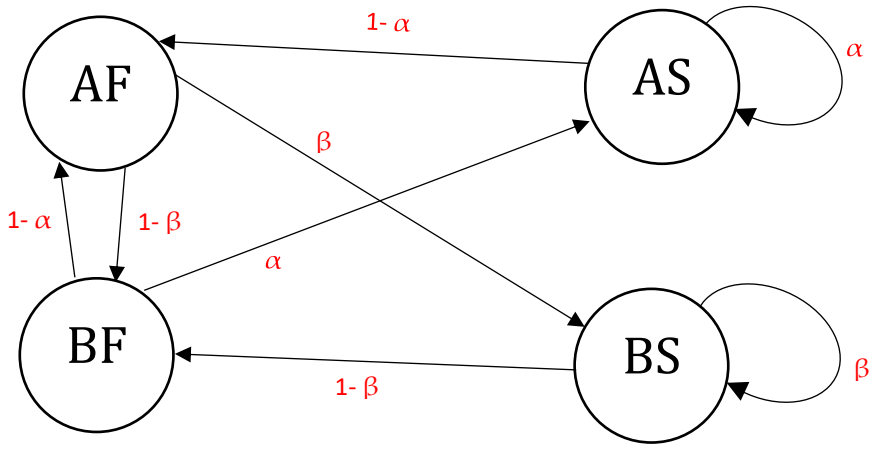
$$P(S) = P(E_1/E_A) \times P(E_A) + P(E_2/E_B) \times P(E_B) = 0.5$$

$$= (0.5 \times \alpha) + (0.5 \times \beta)$$

$$= \left(\frac{1}{2}\right)[\alpha + \beta] = RHS$$

Hence, proved.

(b) State diagram for the Markov Chain



States:

AF: Treatment A fails on n^{th} patient

AS: Treatment A is successful on n^{th} patient

BF: Treatment B fails on n^{th} patient

BS: Treatment B is successful on n^{th} patient

Constructing the transition matrix for the above Markov Chain, we get

x	(AS)	(AF)	(BS)	(BF)
(AS)	α	$1-\alpha$	0	0
(AF)	0	0	β	$1-\beta$
(BS)	0	0	β	$1-\beta$
(BF)	α	$1-\alpha$	0	0

Experimentation and Results for the simulations

The goal was to carry out simulation, testing and verification of the mathematical results we achieved.

Approach

The transition matrix (T) was defined as a 2D matrix. Another matrix π was defined

X	0	1	2	3
0	α	$1-\alpha$	0	0
1	0	0	β	$1-\beta$
2	0	0	β	$1-\beta$
3	α	$1-\alpha$	0	0

$$\pi = \left[\frac{1}{Total\ States}, \frac{1}{Total\ States}, \frac{1}{Total\ States} \ldots\ldots, \frac{1}{Total\ States}\right]$$

For the current example,

$$\pi = [0.25\ 0.25\ 0.25\ 0.25]$$

This matrix was multiplied with the transition matrix (T) and the process was repeated till the time steady state was achieved or a threshold number of iterations (100) had been executed.

$$Finding\ \pi:\ \pi.T = \pi$$

Once, the above state is attained, we can say that equilibrium distribution has been attained.

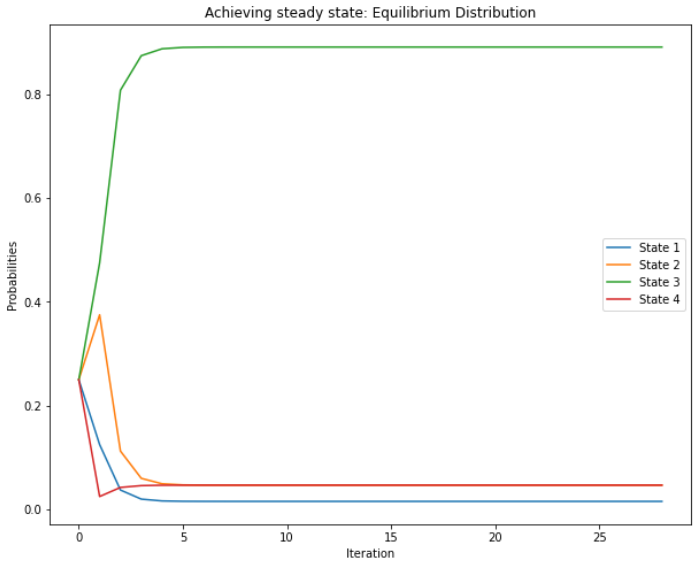


Figure 1: Attaining steady state / equilibrium distribution for the 4 states. The 4 states are State 1 – (AS), State 2 – (AF), State 3 – (BS), State 4 – (BF)

We know $P_r = (\alpha + \beta)/2$. Also, Pt will be the long state probability, or the probability attained after reaching the equilibrium distribution. P_t was calculated by adding the two probabilities for (AS) and (BS), for a large number of α and β.

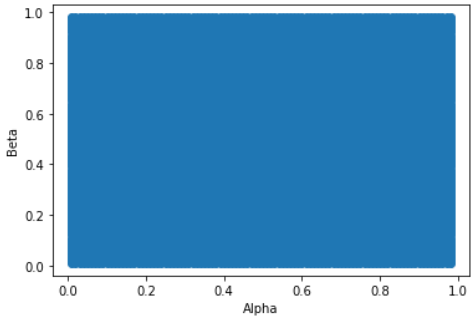


Figure 3: The plot for α and β, covering almost all possible points in the entire space (0,1).

After calculating P_r and P_t for all the above values of α and β, and finding their difference, we obtain the following graph.

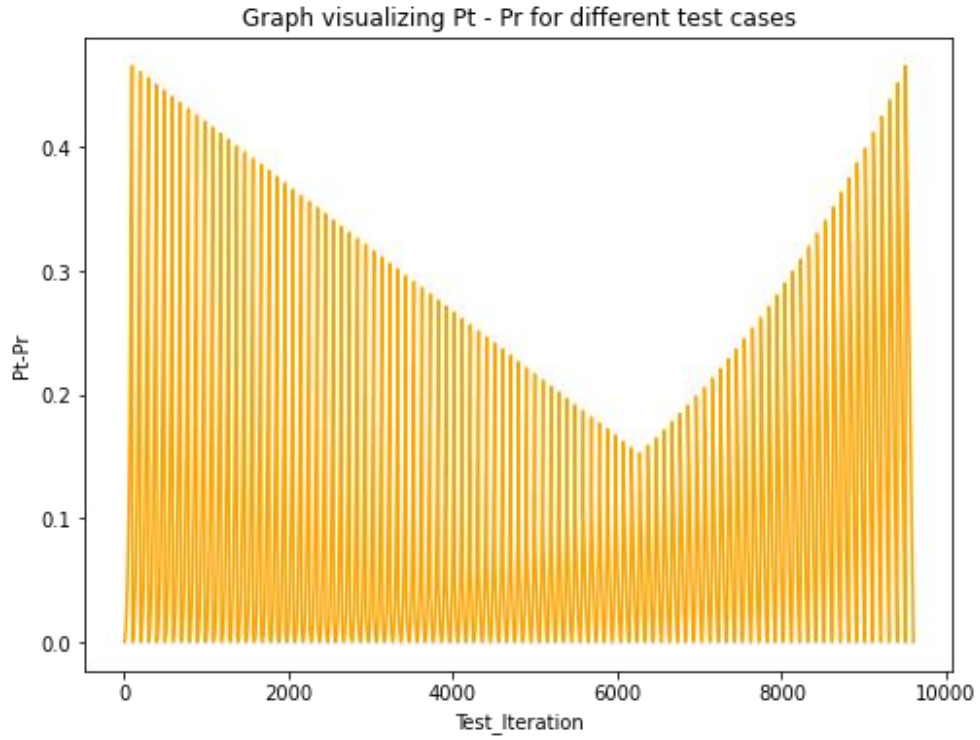


Figure 4: Plot showing the values for $P_t - P_r$ for the different values of α and β

The plot clearly shows that $Pt - Pr \geq 0$ for all α and β. It can be clearly seen that for no value during the Test Iterations the difference goes less than zero. Thus, it can be easily stated that the two-armed bandit strategy is better than the random strategy.

For $\alpha = 0.25$ and $\beta = 0.95$ (arbitrarily chosen)

Pt = 0.906

Pr = 0.6

Long Run transition matrix gives us the Pt for each patient, For $\alpha = 0.25$ and $\beta = 0.95$:

```
[[0.015625 0.046875 0.890625 0.046875]]
```

----------x----------