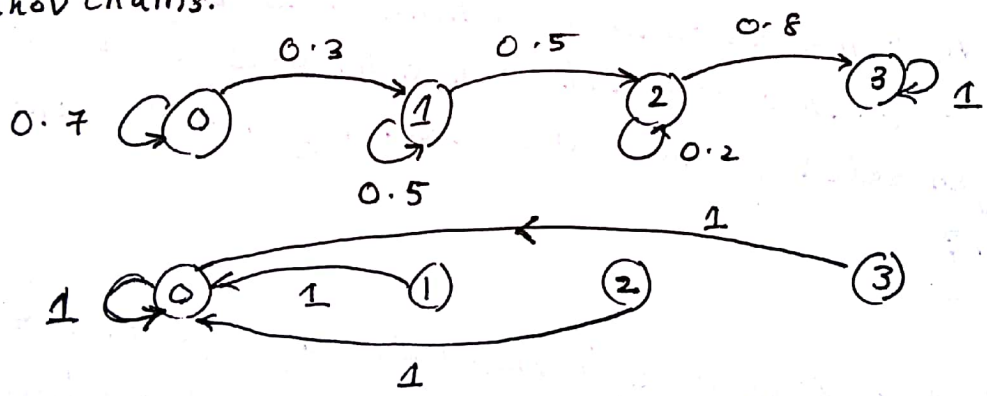
Out : 25/01/2019
Due : 04/02/2019.

AV499 - AVD871 Programming Assignment 3.

In this assignment, you will simulate a system evolution modelled as a Markov chain with actions. At the end of this assignment, you should have the capability to simulate any system (or environment) modelled as a Markov chain and interpret various aspects of the model operationally.

We will simulate a specific example first (recall the machine repair exampl
Suppose the system evolution is modelled in the following way.
The system state denotes the # of failed components in a machine.
We are interested in this state at the start of every day, so that
S_t = # of failed components at the start of day t.
Every day, we can either "run" the system or "repair" the system.
The evolution of S_t under these two actions are given by the following Markov chains.



On any day t, if we decide to run the system, then we obtain a reward which is distributed as an Exponential random variable with mean value = $3 - S_t$. If we decide to repair then we incur a cost (or negative reward) of 100.

1) Suppose the system always starts with state 0, i.e., $S_0 = 0$. Write Python/Matlab code to generate a sample function of the evolution of the system and rewards on every day for 10 days under a control policy that runs the system for the first 5 days, then repairs for the next day, and runs the system for the remaining days.

2) Modify your code so that the control policy is such that on day t whenever the state $S_t \in \{0, 1, 2\}$ the system is run, and if $S_t = 3$, the system is repaired.

3) Recall that a single run of the system evolution yields what is known as a sample function of the system evolution or a realization. Generate a 1000 runs ~~from these thousand runs~~ of the system evolution for the policy in (2).

a) From these 1000 runs or using these 1000 runs can you find out the connection between the conditional probability shown in the transition diagrams and some operational quantity that can be computed from these 1000 runs.
(Hint: how do you interpret conditional probability as the ~~long~~ fraction of time some event has happened).

b) We have said that the reward at a time t is Exponential with mean $= 3 - S_t$. How do you show that the reward sequences that you have generated for these 1000 sample functions follow this distribution?

4*) Can you generalise your code to simulate any system for which the transition probability matrix for each possible action and reward distributions are given.