Homework 8

Problem

Two-Armed Bandit

The purpose of this assignment is to explore conditions under which optimal actions in bandit settings are not discovered even under an optimal Bayesian policy.

You are a decision maker in an infinitely repeated bandit setting with discount factor $gamma$. You have two different arms you can pull. You believe with certainty that arm B pays $rewardB$ dollars per pull. You believe with probability p that arm A pays $rewardA1$ dollars and with probability $1-p$ that it pays $rewardA2$ dollars. (Note that the arms each pay out the same amount per pull: pulling arm A will always result in the same amount, but you are not certain *which* amount will be paid.)

We will give you values of $gamma$, $rewardA1$, $rewardA2$, $rewardB$, p.

Find the value of acting optimally in the Bayesian sense given this prior belief state.