Assignment 3

This assignment is on Friday May 27. Please hand in your assignment to Prof. Bo Wahlberg.

1. Consider a 20- state Markov decision process with state space $X = \{1, 2, \ldots, 20\}$ and action space $U = \{1, 2\}$. The transition matrices $P(u)$ corresponding to the two actions $u = 1$ and $u = 2$, respectively. Construct an example of $P(1)$ and $P(2)$ so that they satisfy the assumptions:

   (A2) $P_i(u) \leq_s P_{i+1}(u)$ (where $\leq_s$ denotes first order stochastic dominance).

   (A4) $P(u)$ is tail supermodular: $\sum_{j \geq l} (P_{xj}(u+1) - P_{xj}(u))$ is increasing in x.

   The costs are $c(x, u) = x + u$.

   (a) Compute the optimal policy via stochastic dynamic programming for a horizon of length 20.

   (b) Prove that the optimal policy is monotone in the state.

   (c) Simulate a sample path of the Markov decision process.

2. Consider the following quickest change detection problem. A sensor has two states: operational (state 2) and failed (state 1). The sensor starts in the operation state and fails at a discrete time instant τ^0 that is geometrically distributed with mean 10.

   (a) Express the dynamics of the sensor state as a two state Markov chain

   (b) Suppose the state of the sensor is observed in zero mean unit variance iid Gaussian noise. Based on these noisy observations, the aim of a decision maker is to detect when the sensor has failed. The decision maker incurs the following costs:

      - If at any time k, the sensor is in the failed state and the decision maker decides that the sensor is operational, the decision maker pays a delay penalty d.
      - If at any time k, the decision maker declares the machine has failed, the problem terminates. If the sensor was in the operational state when the problem teriminates, the decision maker pays a false alarm penalty f. If the sensor was in the failed state when the problem terminates, the decision maker pays no penalty.

   (c) Formulate the above problem as a partially observed Markov decision process. Compute the optimal policy using stochastic dynamic programming. To implement this, you can quantize the belief state space.

   (d) Simulate the performance of the optimal policy on a sample path.