Assume that we have 4 environments with the following reward probability vectors.

\[ En_{0.1} \quad D = (0.7, 0.5, 0.3, 0.2, 0.4, 0.5, 0.4, 0.3, 0.5, 0.2) \]
\[ En_{2.1} \quad D = (0.7, 0.4) \]
\[ En_{2.2} \quad D = (0.6, 0.4) \]
\[ En_{2.3} \quad D = (0.5, 0.4) \]

**Exercise 1.**

A. Implement* the following fixed-structure automata for \( En_{0.1} \).
   - Tsetlin(10,20), Tsetlin(10,40), Tsetlin(10,80)
   - Krinsky(10,20), Krinsky(10,40), Krinsky(10,80)
   - Krylov(10,20), Krylov(10,40), Krinsky(10,80)

B. Plot \( M(n) \) (average received reward up to time \( n \)) and compare the results for them concerning the effect of number of states on the speed of convergence.

**Exercise 2.**

A. Implement the following variable-structure automata for all 4 environments. Set parameters as \( a = 0.01 \) for all automata and \( b = 0.001 \) for LR-eP.
   - LR-I
   - LR-P
   - LR-eP

B. Study the following algorithms from [1] which can be found in the cited URL.
   - TSE Algorithm described in section 2.4.2.2
   - Generalized Pursuit Algorithm introduced in section 4.2

C. Implement the following estimator algorithms for all 4 environments. Set parameter \( \lambda \) as \( \lambda = 0.1 \).
   - CPR-I
   - CPR-P
   - TSE Algorithm
   - Generalized Pursuit Algorithm

D. Describe the results of all the algorithms you implemented in part A and C. For convergence issues, use threshold \( T = 0.99 \) for probability of best action and compare required number of iterations in different algorithms.

**Reference**


* No implementation language is preferred