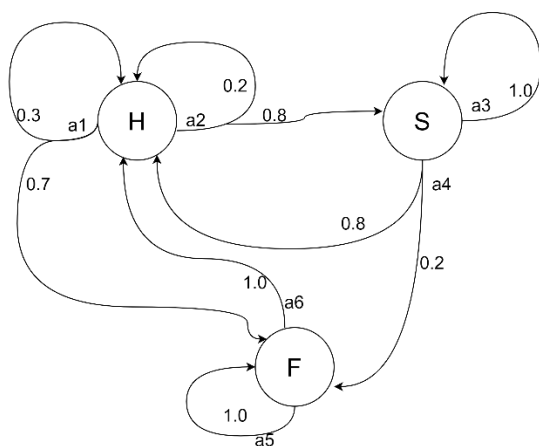


Important: Punctajul pentru exerciții se acordă doar pe baza unei rezolvări detaliate (scrieți mai întâi formulele și apoi înlocuiți cu valorile corespunzătoare)

Subiect 1: Diagrama de mai jos modelează un proces de decizie Markov care surprinde locațiile în care te vei afla de sărbători și tranzițiile dintre acestea. Cele 3 stări sunt: H (acasă), S (la un magazin de cadouri) sau F (la un prieten acasă). Acțiunile sunt: a1 („mergi la un prieten acasă” – cu o mică probabilitate întâlnirea se poate amâna), a2 („cumpără cadou – online sau de la magazin”), a3 („rămâi la magazin”), a4 („mergi de la magazin acasă” – aici accidental te poți întâlni cu un prieten pe stradă și ajungi să mergi la el acasă), a5 („rămâi la prieten acasă”), a6 („mergi de la prieten, acasă la tine”).

Considerați următoarele tranziții care generează recompensele: [(H, a2, S, 1), (S, a3, S, 2), (S, a4, H, -1)]. Valorile inițiale Q sunt setate pe 0, cu excepția: $Q(H, a1) = 1$, $Q(H, a2) = -2$, $Q(S, a3) = 2$, $Q(S, a4) = 1$. Aplicați detaliat pas cu pas ecuația de actualizare a valorilor Q pentru a determina noile valori Q. Rata de învățare este 0.4, iar factorul de discount 0.2.



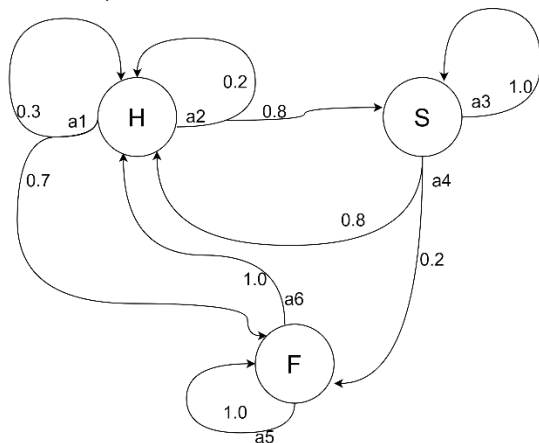
Subiect 2: Ce acțiune vom alege să facem din starea H, conform politicii derivate din Q, pentru următoarea tabelă Q ? Justificați.

	a1	a2	a3	a4	a5	a6
S	-	-	2	1	-	-
H	1	2.5	-	-	-	-
F	-	-	-	-	-1	1

Important: Punctajul pentru exerciții se acordă doar pe baza unei rezolvări detaliate (scrieți mai întâi formulele și apoi înlocuiți cu valorile corespunzătoare)

Subiect 1: Diagrama de mai jos modelează un proces de decizie Markov care surprinde locațiile în care te vei afla de sărbători și tranzițiile dintre acestea. Cele 3 stări sunt: H (acasă), S (la un magazin de cadouri) sau F (la un prieten acasă). Acțiunile sunt: a1 („mergi la un prieten acasă” – cu o mică probabilitate întâlnirea se poate amâna), a2 („cumpără cadou – online sau de la magazin”), a3 („rămâi la magazin”), a4 („mergi de la magazin acasă” – aici accidental te poți întâlni cu un prieten pe stradă și ajungi să mergi la el acasă), a5 („rămâi la prieten acasă”), a6 („mergi de la prieten, acasă la tine”).

Considerați următoarele tranziții care generează recompensele: [(H, a2, H, 3), (H, a1, F, 2), (F, a6, H, 1)]. Valorile inițiale Q sunt setate pe 0, cu excepția: $Q(H, a1) = 2$, $Q(H, a2) = -1$, $Q(F, a5) = 1$, $Q(F, a6) = 2$. Aplicați detaliat pas cu pas ecuația de actualizare a valorilor Q pentru a determina noile valori Q. Rata de învățare este 0.2, iar factorul de discount 0.4.



Subiect 2: Ce acțiune vom alege să facem din starea S, conform politicii derivate din Q, pentru următoarea tabelă Q ? Justificați.

	a1	a2	a3	a4	a5	a6
S	-	-	2	1	-	-
H	3	2.5	-	-	-	-
F	-	-	-	-	-1	1